

Pécsi Tudományegyetem  
Közgazdaságtudományi Kar  
Regionális Politika és Gazdaságtan Doktori Iskola

# **HATÉKONYSÁGJAVÍTÓ ELJÁRÁSOK ASZIMMETRIKUS ELOSZLÁSÚ SOKASÁGBÓL VETT MINTÁKRA A MEZŐGAZDASÁG STATISZTIKÁBAN**

---

**A MAGYARORSZÁGI ÁLLATÁLLOMÁNY PÉLDÁJÁN**

Doktori értekezés

Készítette: Galambosné Tiszberger Mónika

Témavezető:  
Dr. Rappai Gábor  
habil. egyetemi docens

Pécs, 2012.

# Tartalomjegyzék

<b>1. Bevezetés .....</b>	<b>5</b>
1.1. A KUTATÁS CÉLJA, MOTIVÁCIÓI.....	5
1.2. HIPOTÉZISEK ÉS A DOLGOZAT SZERKEZETE .....	7
1.3. A KUTATÁS MÓDSZERTANA .....	10
<b>2. A mezőgazdasági összeírások története és rendszere .....</b>	<b>15</b>
2.1. TÖRTÉNETI ÁTTEKINTÉS .....	15
2.1.1. <i>Cenzusok</i> .....	16
2.1.2. <i>Szőlő- és gyümölcsös-ültetvények összeírása</i> .....	20
2.1.3. <i>Állatállomány összeírások</i> .....	22
2.2. A MAGYAR MEZŐGAZDASÁG TELJESÍTMÉNYÉNEK IDŐBELI ALAKULÁSA .....	27
2.3. A STATISZTIKA JOGI HÁTTÉRÉNEK ALAKULÁSA .....	30
2.4. A MEZŐGAZDASÁGI TERMELÉSSTATISZTIKA RENDSZERE.....	36
2.4.1. <i>A megfigyelési egységek</i> .....	36
2.4.2. <i>Az összeírások rendszere</i> .....	43
<b>3. Mintavétel elmélet.....</b>	<b>49</b>
3.1. FŐBB MINTAVÉTELI MÓDSZEREK .....	50
3.1.1. <i>Egyszerű véletlen mintavétel</i> .....	50
3.1.2. <i>Rétegzett mintavétel</i> .....	52
3.1.3. <i>Csoportos mintavétel</i> .....	59
3.1.4. <i>Többlépcsős mintavétel</i> .....	62
3.1.5. <i>Mintavételi módszerek csoportosítása</i> .....	63
3.1.6. <i>Nem véletlenül alapuló mintavételi módszerek</i> .....	64
3.2. MINTAVÉTEL 2000 UTÁN A MAGYAR MEZŐGAZDASÁG STATISZTIKÁBAN .....	65
3.2.1. <i>2000-2005</i> .....	65
3.2.2. <i>2006-2010</i> .....	68
3.2.3. <i>2010 után</i> .....	70
3.3. A STATISZTIKAI ADATOK MINŐSÉGE .....	72
<b>4. A szimulációs számítások háttére, az adatállomány jellemzői .....</b>	<b>78</b>
4.1. A KÍSÉRLETI ADATBÁZIS .....	78
4.2. AZ ÁLLATTARTÁS SPECIÁLIS JELLEMZŐI.....	80

4.3. AZ ASZIMMETRIA MÉRÉSE A MEZŐGAZDASÁGBAN .....	84
4.4. LEÍRÓ ELEMZÉS ÁLLATFAJONKÉNT .....	93
4.5. ALKALMAZOTT MINTAVÉTELI MÓDSZEREK .....	100
<b>5. Szimulációs eredmények .....</b>	<b>108</b>
5.1. EGYSZERŰ VÉLETLEN MINTAVÉTEL .....	108
5.2. RÉTEGZETT MINTAVÉTEL.....	111
5.2.1. Rétegzés régiók szerint.....	115
5.2.2. Módszertani megfontolás .....	124
5.2.3. Rétegzés nagyságkategóriák szerint.....	127
5.2.4. További lehetőségek az OPS meghatározásához .....	144
5.2.5. Rétegzés régiók és nagyságkategóriák szerint .....	153
5.3. MÓDSZERTANI MEGFONTOLÁSOK A KIVÁLASZTÁSI ARÁNNYAL KAPCSOLATBAN .....	156
5.4. KÍSÉRLETI EREDMÉNYEK A 2010-ES ÁLTALÁNOS MEZŐGAZDASÁGI ÖSSZEÍRÁSBÓL .....	160
<b>6. Összegzés .....</b>	<b>164</b>
<b>Ábrák jegyzéke.....</b>	<b>170</b>
<b>Táblázatok jegyzéke.....</b>	<b>171</b>
<b>Irodalomjegyzék.....</b>	<b>173</b>
<b>Mellékletek.....</b>	<b>181</b>
<b>Publikációs jegyzék.....</b>	<b>196</b>

## **Rövidítések jegyzéke**

ÁMÖ: Általános Mezőgazdasági Összeírás

DEFF: design effect, tervhatásossági mutató

EU: Európai Unió

EV: egyszerű véletlen

FADN: Mezőgazdasági Számviteli Információs Hálózat (Farm Accountancy Data Network)

GSZÖ: Gazdaságszerkezeti Összeírás

KSH: Központi Statisztikai Hivatal

NO: Neyman-féle optimális allokáció rétegzett mintavételnél

OPS: optimális réteghatár (optimum point of stratification)

OSAP: Országos Statisztikai Adatgyűjtési Program

OST: Országos Statisztikai Tanács

SEFF: size efficiency, mérethatásossági mutató

SFH: Standard Fedezeti Hozzájárulás (Standard Gross Margin)

TATS: take all – take some mintavételi módszer

## 1. Bevezetés

A statisztikai vizsgálatok minden esetben az igények és a lehetőségek közötti szakadék áthidalását biztosító kompromisszumok mentén valósulnak meg. A vizsgálat módszerének, tárgyának, a kutatás mélységének és méretének meghatározása a hivatalos statisztikában csak részben a statisztikus feladata. A mezőgazdaság statisztikában a jelenleg hatályos jogszabályi háttér, az Eurostat elvárásai az összeírások kereteit és rendszerét szinte minden oldalról körülfonják. Nómenklatúrák, definíciók, határidők, lefedettség, mintavételi hiba, stb. mind eldöntött kérdések. A statisztikus munkájában tehát nem az előbbiekre vonatkozó kérdések megválaszolása jelenti a nehézséget, hanem az előírásoknak való megfelelés módjának a megtalálása. A keretekről nem kell gondolkodni. Sokkal inkább az ezekhez igazodó mintavétel módja és mikéntje jelenti az izgalmas kérdéseket. A dolgozat egy ilyen módszerkeresésnek a menetét mutatja be a keretek ismertetése mellett, a mezőgazdaságot leíró sokaságokra jellemző aszimmetrikus eloszlások mellett.

### *1.1. A kutatás célja, motivációi*

A magyar mezőgazdaság statisztikájának a Központi Statisztikai Hivatalban (KSH) 150 éves hagyományai vannak. A mezőgazdaságról gyűjtött adatok – mivel az élelmiszerellátást meghatározó ágazat, és a lakosság viszonylag nagy részének jelentett, és háztáji szinten még ma is munkalehetőséget jelent – mindig is nagy érdeklődésre tartottak számot. Korábban leginkább hazai felhasználók, ma már a nemzetközi összehasonlíthatóság és az Európai Unió jogszabályai befolyásolják a legnagyobb mértékben az adatgyűjtések kereteit. Az összeírások pontosságára (a mintavételből eredő standard hibák nagyságára) vonatkozó előírások meglehetősen szigorúak, és egyre inkább jogszabályi szinten szabályozottak. Ezek a követelmények nehéz helyzetet teremtenek a statisztikusok számára, hiszen a magyar mezőgazdaság szerkezeti sajátosságai a legtöbb változó szempontjából speciális, extrém tulajdonságokkal bírnak. (A mezőgazdaság sokszínűsége éppen ezért a gazdasági szintű hatékonyságméréseket is megnehezíti. [Kerékgyártóné-Mundruczó 1976]) Mindemellett, noha ez módszertani szempontból nem szerencsés, költséghatékonysági okból a mezőgazdaság statisztika adatgyűjtéseinek jellemzője a többcélúság. Ez annyit jelent, hogy egy-egy kérdőív több, különböző természetű változót és esetleg több témát is felölel. (Kish

1989) A reprezentatív mintának pedig lehetőség szerint minden szempontot egyidejűleg kellene figyelembe vennie, ami csak számos kompromisszum mellett lehetséges.

A KSH-nál töltött közel 5 év alatt lehetőségem nyílt részletesen megismerni a magyar mezőgazdaság aktuális helyzetét, a gazdaságok szerkezetét, sajátosságait és az adat-előállítás folyamatát, részleteit. A mintavételi és becslési módszerek kidolgozása terén azonban meglehetősen kötött pályán kellett mozogni a szervezeti és anyagi feltételek miatt. Az extrém körülmények mellett ugyanakkor éppen az elméleti lehetőségek minél nagyobb mértékű kiaknázása hozhat hatékony megoldási javaslatokat.

A magyar mezőgazdaság szereplőinek két nagy körét különítjük el. Egyiket a gazdasági szervezetek alkotják, akik számosságukat tekintve nem képviselnek jelentős nagyságrendet (viszonylag stabilan 8000 körül mozog a számuk), viszont a mezőgazdasági termelés értékének nagyjából a felét adják. Az ő megfigyelésük teljes körűen történik, ezért mintavétel-módszertani szempontból nem okoznak fejtörést, mintavételi hiba nélkül kerülnek összeírásra. A másik kör az egyéni gazdaságok. A termelési érték másik felét hozzák létre évről-évre, de számosságuk nagyságrendekkel magasabb, és gyorsabb ütemben változó. 2007-ben még több, mint 600 ezren voltak, a legfrissebb, 2010-es census szerint pedig 566 ezer a számosságuk. A mintavételes eljárásoknak ezen gazdálkodói kör tagjai képezik a megfigyelési egységeit, hiszen elképzelhetetlen, hogy egy ekkora tömeget évente többször is, teljes egészében megfigyeljünk. Ők jelentik tehát a célsokaságot a vizsgálódások szempontjából.

A mezőgazdasági termelésstatisztika két nagy témája a földterület és az állatállomány megfigyelése, amelyek jellemzői, természete, a termelésben betöltött szerepe meglehetősen eltérő. A földterületnek megvan az a sajátossága, hogy a hasznosítás módja viszonylag lassan változik, valamint az ország teljes területe pontosan meghatározza a nagyságát. A cél itt a művelési ágak rövid távon viszonylag állandó megoszlásán belül, az egyes növények termőterületének és termésmennyiségének a meghatározása. A jelenlegi rendszerben ezt a szervezetek teljes körű megfigyelésén, az egyéni gazdaságok mintavételes adataiból származó teljeskörűsített adatokon túl, regionális, illetve megyei szinten úgynevezett szakértői becslések is kiegészítik. Ez utóbbi helyi ismeretekre, különböző adminisztratív forrásokra (például terméktanácsok adatai, helyközségek információi) támaszkodva pontosítja a végleges számokat. Az állatállomány esetében a változékonyság akár rövidebb idő alatt is jelentős. Az állattartás jellegéből adódóan könnyebb a gazdaságoknak is rugalmasan igazodni a változó feltételekhez, gazdasági környezethez mint az a földterület esetében. A két terület

megfigyelése, számbavétele éppen ezért eltérő követelményeket támaszt. Jelen dolgozat nem fedi le a teljes termelésstatisztikát, hanem csak az állatállomány mintavételes összeírásának lehetőségeit vizsgálja a hatályos Európai Unió jogszabályoknak való megfelelés tükrében. Az EU-s előírásoknak megfelelően 3 állatfaj van az elemzések fókuszában: a szarvasmarha, a sertés és a juh. Mindhárom állatfaj eloszlása (a gazdaságok eloszlása a tartott állatszámok szerint) közepesnek, illetve erősnek nevezhető, jobb oldali aszimmetriát mutat. Ez jelentős terjedelem és szóródás mellett valósul meg. A kiinduló feltételek tehát jelentős kihívásokat támasztanak a mintavételezés mikéntjét meghatározni próbáló statisztikusok elé.

A dolgozat célja alapvetően kettős. Van egyrészt egy általánosabb, elméleti jellegű cél, másrészt egy konkrét gyakorlati megoldás kidolgozására irányuló motiváció. Leegyszerűsítve a két alapmotívum az alábbi:

- Általánosan hasznosítható megoldások keresése és kidolgozása jelentős aszimmetriát mutató eloszlással rendelkező változók esetén a hatékony mintavételi terv kidolgozásához.
- Az állatállomány becslések jogszabályban meghatározott mintavételi hibáinak teljesítéséhez hatékony mintavételi terv kidolgozása, természetesen az elméleti módszerek, eljárások alkalmazásával.

A kutató munkát végig áthatja ez a kettősség, tehát az elméleti lehetőségek és a gyakorlati megvalósíthatóság ütköztetése. Így tehát az elméleti áttekintés nem öncélú, az empirikus adatok elemzése, vizsgálata pedig inspirálja a módszertan változatos kihasználását. (Hunyadi-Rappai 1999)

## ***1.2. Hipotézisek és a dolgozat szerkezete***

Az előzetes ismeretek és tapasztalatok alapján már a dolgozat írásának kezdetén megfogalmazódott bennem néhány feltevés, és ahogy az lenni szokott, a kutatás során is

formálódtak ezek az elképzelések, újabb észrevételekkel kiegészülve. 6 hipotézist fogok megvizsgálni, illetve „tesztelni”<sup>1</sup> a dolgozatban:

#### 1. hipotézis (H1):

Sok kritika éri a mezőgazdaság statisztikát amiatt, hogy különböző témájú rendszeres összeírásokhoz ugyanazt az „univerzális” mintasokaságot figyelik meg. Feltételezhető ugyanis, hogy témánként csak a potenciális gazdaságokat alapsokaságnak tekintve hatékonyabban lehetne mintavételi tervet kidolgozni, és ezzel növelhető lenne az adatok minősége, illetve csökkenthető lenne az azonos minőség mellett szükséges mintavételi arány. Mivel a dolgozat középpontjában az állatállomány példája áll, ezért előzetesen azt feltételeztem, hogy **az állatállomány megfigyeléseket hatékonyabban lehet végrehajtani egy szűkített alapsokasági szemlélet mellett, vagyis alacsonyabb mintavételi arányokra van szükség, ha csak az állattartó gazdaságokból indulunk ki, a magasabb homogenitás miatt.** → T5<sup>2</sup>

#### 2. hipotézis (H2):

A **területi reprezentativitás** igénye hazai és nemzetközi oldalról egyaránt jelentkezik. Ez előre kijelöl egyfajta rétegzési szempontot, ami a régiók szerinti csoportosítás. Mivel Magyarországra – területének méretéből, az egységesnek mondható klimatikus és területi viszonyokból adódóan – nem jellemző, hogy a régiók nagyon eltérő képet mutassanak a mezőgazdasági-termelési sajátosságokban, ezért a **régiós rétegzés** önmagában szinte biztos, hogy **nem javítja a mintavétel hatékonyságát**. Ezt másképpen is megfogalmazhatjuk: nincsen szignifikáns kapcsolat a régió és a vizsgált ismérvek között. Egy harmadik megfogalmazásban pedig azt is mondhatjuk, hogy az országos szinten jelentkező összetétel (heterogenitás) megmarad a régiók szintjén is, mintegy „kicsiben” leképezve azt. Megint másként fogalmazva: a régiós rétegzést az Európai Unió jogszabályok előírják (emellett

---

<sup>1</sup> A tesztelés azért került idézőjelbe, mert statisztikai értelemben a hipotéziseket megfelelő statisztikai próbával, próbafüggvénnyel szokás ellenőrizni, de itt természetesen nem erről lesz szó a dolgozat és a téma jellegéből adódóan.

<sup>2</sup> A dolgozat szerkezete nem követi a fenti hipotézisek menetét. A hipotézisek sorrendjét a feltevések logikája adja, ugyanakkor a bizonyítás több esetben összefügg, illetve több lépésből áll, ezért a hipotézisek tárgyalásának sorrendje nem követi a sorszámokat. A könnyebb eligazodás kedvéért került oda a hipotézisek végére a hozzájuk kapcsolódó tétel sorszáma, Tx jelöléssel.



természetesen ez hazai felhasználói igény is), viszont mintavétel-technikai szempontból ez szükségtelen lenne. → T3

### 3. hipotézis (H3):

A 2. hipotézis kapcsán kezdtem el gondolkodni a rétegzés sajátosságain. Míg a köztudatban (és a tankönyvekben) elterjedt vélekedés szerint a rétegzett mintavétel egy hatékonyság-javító eljárás, addig a korábbi tapasztalataim alapján azt látom, hogy **szignifikáns hatékonyságnövekedés csak bizonyos – megfelelően kiválasztott – réteggépző ismérvek alkalmazása esetén érhető el**. Ugyanakkor **rétegzést nem feltétlenül és nem kizárólag azért alkalmazunk a gyakorlatban, hogy növeljük a becslés hatékonyságát**, hanem más motivációk is állhatnak a háttérben. Mindezt felteszem, hogy lenne lehetőség a **rétegzett mintavétel alcsoportokra osztására** annak mentén, hogy az ténylegesen javítja-e a mintavétel hatékonyságát. → T1/A, T1/B

### 4. hipotézis (H4):

Mivel a **jogszabályi előírásokban a becslések mintavételi hibájának maximuma előre adott**, ezért már a vizsgálódások legelején világossá vált, hogy a „szokásos” módon nem lesz célravezető a különböző mintavételi tervek hatékonyságának összehasonlítása. Hiszen két mintavételi tervet úgy szokás összehasonlítani, hogy azonos mintanagyság mellett vizsgáljuk meg, hogy a becslés hibája (a becslőfüggvény varianciája) hogyan alakul. Az ún. **tervhatásossági mutató** tehát **ebben a helyzetben nem célravezető**. Felteszem tehát, hogy a gyakorlati munkához **kialakítható ennél megfelelőbb mutató két, azonos standard hibát eredményező mintavételi terv hatékonyságának összehasonlítására**. → T2

### 5. hipotézis (H5):

A jelentős aszimmetriával terhelt sokaságokra alkalmazhatónak véltem egy olyan módszert, ami a viszonylag magas értékekkel rendelkező gazdaságokat teljes körűen megfigyeli, és igazi mintavételi tervet csak a kisebbekre dolgoz ki. Ez tulajdonképpen egyfajta rétegzés, a rétegek közötti speciális allokációval, amit a szakirodalom **„take all – take some”** (TATS) módszernek nevez. Feltételezem, hogy empirikus adatok alapján, illetve elméleti szempontból is meghatározható, hogy milyen feltételek mellett lehet ez a módszer hatékonyabb mintavételi eljárás a **Neyman-féle optimális elosztással** szemben. A hipotézis

tehát abban áll, hogy **az aszimmetria erőssége és a szórás mértéke meghatározza, hogy melyik módszer a hatékonyabb.** → T6

#### 6. hipotézis (H6):

A korábbi tapasztalatok és ismeretek azt mutatták, hogy a három – a jogszabályokban kitüntetett figyelmet élvező – állatfaj közül a juhok jelentik majd a legnagyobb problémát. Egyszerűen azért, mert a szóródás itt a legmagasabb, és itt vannak a legtávolabb a magas értékek az átlagtól, vagyis a legmagasabb outlier-eket, kiugró értékeket a juhoknál találjuk. Azt feltételeztem, hogy **a sertések és a szarvasmarhák esetében nem lesz annyira nehéz alacsony mintavételi arányokat produkálni a megfelelő rétegzés mellett, és éppen ezért ennél a két fajnál nem kell a gyakorlatban sem annyira szigorú és pontos rétegzési szabályokat alkalmazni.** Ferde eloszlás jelentkezik ezeknél is, de nem olyan mértékben, mint a juhok esetében. → T4

A mezőgazdaság statisztika történeti áttekintése, a magyar mezőgazdasági ágazat időbeli alakulása és szerkezetének bemutatása valamint a statisztika jogszabályi háttere adják a téma keretét az empiria oldaláról. A mintavételi módszerek elméleti áttekintése jelentik a bázist a későbbi szimulációs számítások megalapozásához. A kísérleti adatbázis bemutatása, az aszimmetria mérésének sajátosságai, az állattartás speciális jellemzői és a KSH-ban alkalmazott mintavételi módszerek jellemzése, összefoglalása vezeti fel a későbbi eredményeket. A dolgozat magját a szimulációs számítások jelentik. A különböző mintavételi módszerek, illetve azok kombinálásából származó eredmények jelenthetnek használható támpontot a mezőgazdaság statisztikai gyakorlat számára. A dolgozatot az eredmények összefoglalása mellett a további kutatási irányok felvázolása zárja.

### ***1.3. A kutatás módszertana***

A kutatás kiegészítésének, illetve a téma keretbe foglalásának, a háttér kidolgozásának igénye tette szükségessé a magyar mezőgazdaság statisztika történeti áttekintését, az EU-s adatgyűjtési, adatszolgáltatási szabályok, a jelenlegi termelésstatisztika rendszerének bemutatását. Ugyanilyen motívumok mentén került a dolgozatba a statisztai adatok minőségével foglalkozó alfejezet is. Ezek inkább leíró, a témák legfontosabb elemeit kiemelő

részei a dolgozatnak. Csakúgy, mint a későbbi számításokhoz már jobban kötődő, a mintavételi módszereket áttekintő oldalak.

A kutatási módszertan egyik része természetesen a hazai és nemzetközi szakirodalom feldolgozása volt. (Bertáné Németh 2001) Az általános statisztikai szakkönyvek mellett elsősorban az extrém ferde, illetve konkrétan mezőgazdasági vonatkozású könyveket és publikációkat kerestem és használtam fel. Azt tapasztaltam, hogy a – talán speciális – téma nem igazán része napjaink aktuális szakirodalmának. A külföldi szakirodalomban is a legtöbb kapcsolódó cikket és tanulmányt az 1950-es és 1960-as évekből találtam. A téma, aktualitása ellenére, nem kap akkora hangsúlyt a frissebb szakirodalomban.

Mivel aszimmetrikus eloszlásokkal foglalkozik a dolgozat, ezért elméleti szempontból érdekes és fontos volt az aszimmetria mérésének vizsgálata is. A szakirodalomban ez még napjainkban is viszonylag aktuális téma. Ugyanakkor érdekes, hogy a statisztika tudomány mellett az orvostudomány és a biológia kutatói is sokat foglalkoznak ezzel a kérdéssel.

A kutatás empirikus részéhez a legfontosabb alapot a Központi Statisztikai Hivatal 2007-es Gazdaságszerkezeti Összeírásának (GSZÖ 2007) adatbázisa adta. A Hivatal engedélyével a kutatás céljaira felhasználhattam a GSZÖ 2007 teljes adatbázisát, természetesen az adatvédelmi előírások betartása mellett. Kaptam egy ideiglenes felhasználónevet, és a KSH Pécsi Igazgatóságán lehetőséget és számítógépet ahhoz, hogy elvégezzem a szükséges lekérdezéseket. Az összeírás közel 111 ezer egyéni gazdaságot ért el, vagyis ez az az elemszám, amivel dolgozni tudtam. Egy ekkora állomány azért nagyon jó kiindulási alap, mert a 2007-es magyarországi egyéni gazdaságok számának 18 százalékát teszi ki. Mivel ez viszonylag magas arány (a mintavételi arány általában néhány százalék), és számosságát tekintve is nagynak mondható az adatbázis, megfelelő alapnak tartom a számítások elvégzéséhez.

Néhány összefüggés kapcsán fontos volt időben is megnézni az adatok alakulását. Ehhez elsősorban a KSH honlapján, illetve a kiadványokban található adatokra támaszkodtam. Részletesebben felhasználtam a 2003-as és a 2005-ös gazdaságszerkezeti összeírások adatait is.

A KSH-s adatbázisok Oracle alapon szerveződnek. A lekérdezésekhez az SQL nyelv felhasználására, és a KSH-ban használatos névkonvenciók, szabványok, struktúrák ismeretére volt szükség. Szerencsére ezekkel még a Hivatali évek alatt megismerkedtem, így nem

okozott problémát a lekérdezések elvégzése. A lekérdezések során körülbelül 10 adatbázis táblát használtam fel, amelyek egy része a konkrét mezőgazdasági adatokat tartalmazta. Másik részük pedig inkább technikai jellegű, hiszen a felhasznált nomenklatúrák fordítókulcsaira, táblák összekapcsolására is szükség volt.

Szimulációnak nevezem a kutatás során használt módszert, hiszen egy – viszonylag nagy – mintasokaság adatait tekintettem alapsokaságnak (mintegy szimulálva az alapsokaság ismeretét), és így az általában ismeretlen sokasági szórásokat ismertnek feltételeztem. Ezek segítségével vizsgáltam meg a három kiemelt állatfaj előírt pontosságú becsléséhez szükséges mintaelemszámok alakulását, különböző mintavételi módszerek mellett. Az egyes módszerek hatékonyságbeli különbségeit is vizsgálom, és választ keresek a különbségek okaira.

A sokaság bemutatásához, megismeréséhez a három állatfajra (szarvasmarha, sertés, juh) vonatkozó leíró statisztikai elemzések adják meg az alapot.

A rétegző ismérvek vizsgálatánál a kapcsolatok létének teszteléséhez függetlenségvizsgálatot (Khi-négyzet próba), a kapcsolatok erősségének méréséhez pedig kapcsolatszorossági mérőszámokat (Cramer együttható, szóráshányados) is felhasználtam.

A szöveg és a számítások illusztrálása, illetve az eredmények összefoglalása céljából ábrák és táblázatok teszik változatosabbá a dolgozatot. Ezek szerkesztése közben szem előtt tartottam az egyszerűség, pontosság és célszerűség szempontjait. (Hunyadi, 2002)

A szükséges számításokat alapvetően az MS Excel szoftver alkalmazásával végeztem el. (Rappai 2001) Az ábrák zöme is innen származik. Az SPSS súlyozási lehetőségét felhasználva készültek a gyakorisági sorokból a leíró statisztikai eredmények. (Székelyi-Barna 2002)

A dolgozatot kettősség jellemzi abból a szempontból is, hogy egyrészt a statisztika elméleti háttere motiválta a számításokat, másrészt viszont a kapott eredmények adtak ötletet az elméleti továbbgondolásra.

## Köszönetnyilvánítás

Végezetül itt szeretném megragadni az alkalmat, hogy köszönetet mondjak mindazoknak, akik segítettek, motiválták, támogatták a kutatói folyamatot és a dolgozat megírását. Elsőként köszönöm a Regionális Politika és Gazdaságtan Doktori Iskolának, a doktori iskola akkori és jelenlegi vezetőjének, Buday-Sántha Attilának és Varga Attilának és minden oktatónak a motiválást, az átadott tudást. Az abszolutóriummi dolgozat, és a műhelyvita opponenseinek, bírálóinak, Hanyecz Lajosnak, Laczka Évának és Mellár Tamásnak, akik részletes kritikai észrevételeikkel nagyban hozzájárultak a végső változatának még tartalmasabbá tételéhez. A kutatás részeredményeit tárgyaló, a Statisztikai Szemlében megjelent cikkem lektorainak, Hunyadi Lászlónak és Vita Lászlónak is hálás vagyok a hasznos tanácsokért, észrevételekért. Héráné Tóth Andreának és Márk Saroltának köszönöm a mindenre kiterjedő ügyintézkést.

Nagyon megkönnyíti a kutató munkáját, ha a közvetlen intézeti munkatársai, kollégái inspirálják, motiválják, és ötleteikkel előrébb viszik a kutatást. Szeretném megköszönni a tanszéki kollégáknak, Kehl Dánielnek, Herman Sándornak és Sipos Bélának a folyamatos érdeklődést és támogatást. Daninak külön köszönöm a sok cikket, amiknek a felkutatását helyettem végezte el.

A dolgozat háttere, a kutatási alapötlet a Központi Statisztikai Hivatalban töltött éveknek köszönhetően alakult ki. Nemcsak az ott eltöltött 5 év alatt, de azóta is folyamatosan számíthattam a kollégák együttműködésére, segítségére. Nekik is köszönök mindent. Külön kiemelném Pintér Lászlót, aki módszertani kérdésekben mindig a biztos támaszt jelentette, Tóth Pétert, aki minden állatállománnyal kapcsolatos kérdésre válaszolt, Tohai Lászlót és Trucza Istvánt, akiket a lekérdezések megírása közben bármikor felhívhattam, és Kincses Áront, aki a matematikus szemével nézte a készülő dolgozatot. A KSH Pécsi Igazgatóságának is köszönöm, hogy helyet és lehetőséget biztosított az adatbázishoz való hozzáféréshez, ami nélkül a számításokhoz szükséges adatokat nem tudtam volna összegyűjteni.

Számomra a családom segítsége is döntő jelentőséggel bír, hiszen a kettő, illetve három gyermek mellett csak úgy lehetett haladni a munkával, ha a család ehhez időt biztosított. Ezért köszönöm Apának, a Mamiknak, a Papának, Editnek, Áginak a gyerekek szórakoztatását és felügyeletét, amíg én a gép előtt ültem, és hogy kétségek nélkül hittek bennem. Külön köszönöm férjemnek az olvasószerkesztést, és az állandó, de kedves „noszogatást”, ami bizonyára hónapokkal előrébb hozta a disszertáció befejezését.

Végül, pedig témavezetőmnek, Rappai Gábornak mondok köszönetet. Az a látásmód és szemlélet rengeteget adott nekem a dolgozat megírása közben, és a kutatáshoz való hozzáállásban is, amit tőle kaptam. Minden válasza túlmutatott a kérdéseimen, és újabb ötleteket generált. Nagyon sokat tanultam tőle, és mindig számíthattam, támaszkodhattam a segítségére.

## **2. A mezőgazdasági összeírások története és rendszere**

A mezőgazdaság statisztika működését, rendszerét a kialakulásán és a történetén keresztül érthetjük meg a legjobban. A dolgozatban a mintavételi lehetőségekre koncentrálok, azonban az összeírások mintegy 150 éves tapasztalata elengedhetetlen a probléma teljes körű megértéséhez. A censusok (teljes körű összeírások) és egyéb adatgyűjtések történetén túl a statisztikai munka keretét meghatározó jogi szabályozással is foglalkozik ez a nyitó fejezet. Mindez azért is szükséges, mert a későbbi részekben, a konkrét mintavételi problémák felvetésekor a múltban gyökerező okokra fogok majd visszautalni.

### ***2.1. Történeti áttekintés***

Már az osztrák statisztikusok is foglalkoztak a magyar viszonyok leírásával, de a statisztikai adatok államigazgatási célú feljegyzésekből kerültek ki, ami akkor nagyon fontos lépés volt, de minőségében meg sem közelítette a későbbi magyar hivatalos statisztikát.

Egy országos hatáskörű statisztikai szervezet létrehozását már a magyar szabadságharc kirobbanása előtt (1848-ban) megkísérelték. Fényes Elek vezetésével egy viszonylag kis létszámú, Országos Statisztikai Hivatal névre hallgató szervezeti egység jött létre a Belügyminisztérium keretei között. Ez a kísérlet történelmi jelentőségű ugyan, de rövid életűnek bizonyult, és a szabadságharc bukása után meg is szűnt a hivatal. A kiegyezést követően a „Földművelés-Ipar és Kereskedelmi Minisztérium” irányítása alatt alakult meg az „igazi” hivatal, Keleti Károly vezetésével, 1867. május 25-én. Ezt a dátumot tekintjük a hivatalos statisztikai szolgálat születésnapjának. A Keleti Károly nevéhez fűződő módszerek a korábbi „leíróbb” jellegű eredményekhez képest merőben új irányt mutattak a hivatalos statisztika fejlődésében.

Magyarországon a mezőgazdasági statisztikának történelmi hagyományai vannak. Az 1800-as évektől folytak összeírások a mezőgazdaság egy-egy részterületére vonatkozóan. Az első állatszamlálást már 1884-ben végrehajtották. A censusok sorát az 1895-ös nyitotta meg, ami mérföldkőnek számít a mezőgazdasági statisztika történetében. Ettől kezdve, kisebb-nagyobb időközönként rendszeressé váltak a teljes körű összeírások. Emellett számos tematikus adatgyűjtésre is sor került: gyümölcsfa, gép, szőlő megfigyelésekre. 1970-ben

csatlakozott Magyarország a FAO<sup>3</sup> (Egyesült Nemzetek Élelmezési és Mezőgazdasági Világszervezete – Food and Agriculture Organization of the United Nations) világcenzusának ajánlásaihoz. Ettől kezdve 10 évente zajlanak az alapsokaságra kiterjedő adatgyűjtések (alapsokaság a mezőgazdasági tevékenységet folytató gazdasági szervezetek, és a bizonyos fizikai küszöbértéket meghaladó egyéni gazdaságok köre, lásd 2.4.1 fejezet).

A rendszerváltást követő első teljes körű összeírást a Hivatal 1991-ben hajtotta végre. A cenzust 1994-ben gazdaságszerkezeti felvétel követte, ami csak szűk mutatókörre vonatkozott és a lefedettsége sem volt teljes. A teljes körű megfigyelések sorában a 2000. évi Általános Mezőgazdasági Összeírás (ÁMÖ 2000) történelmi jelentőségű, mivel ez az első olyan teljes körű adat-felvételezés, amely a FAO igényeken túl szem előtt tartotta az Európai Unió (EU) vonatkozó előírásait is. Ezt azért is fontos hangsúlyozni, mert a harmonizációs munka a statisztikában meglehetősen nehéz feladat, és sok ország küzd a megoldásával. (David 1998) Az ÁMÖ 2000 óta 3 gazdaságszerkezeti összeírássra (GSZÖ) került sor (2003-ban, 2005-ben és 2007-ben). Ezek mindegyike megfelelt az Európai Unió statisztikai hivatala (Eurostat) által támasztott módszertani és tartalmi előírásoknak. A nagyobb lélegzetű összeírások közötti időben rendszeres éves, évközi adatgyűjtésekkel biztosítja a Központi Statisztikai Hivatal (KSH) a mezőgazdasági adatok begyűjtését, előállítását. Jelenleg a 2010. évi cenzus adatainak publikálási munkálatai zajlanak.

### 2.1.1. *Cenzusok*

Az első részletes valamennyi önálló mezőgazdasági egységre kiterjedő összeírást az 1895. évben hajtották végre, november 20-ai eszmei időponttal. Legnagyobb érdeme az első cenzusnak az az átgondolt, jól strukturált mutatókör, ami az agrárium akkori helyzetét átfogóan leírta. A Keleti Károly vezetése alatt kidolgozott kérdőív szerkesztésekor alkalmazott minőségi szempontok napjainkban is megállják a helyüket. A részletességen és a nagy volumenű munkán túl fontos eredmény az is, hogy a kor feltételei mellett 20 hónappal

---

<sup>3</sup> Két magyar szakember is dolgozott magas pozícióban a FAO szervezetében. Párniczky Gábor 1983 és 1986 között volt a Statisztikai Elemzési Osztály vezetője. Őt 6 évre Ay János követte ugyanitt, aki aztán 1992 és 1994 között a Statisztikai Igazgatóság vezetőjeként dolgozott. (Gömbös, 2008)



az eszmei időpont után már közzétették a legfontosabb adatokat községsorosan. Az 1. ábra ennek a kiadványnak a címlapját mutatja. A publikáció előszavát Dr. Darányi Ignác (földmivelésügyi m. kir. Minister) írta. A munka szellemiségének illusztrálásához csak egy rövid bekezdést idéznék ebből:

*„A magyar agrár-statisztika kiépítése, melyet hivatali elődeim oly nagy buzgalommal és szakavatottsággal indítottak meg, ezzel a tényleges megvalósulás stádiumába lépett. Gondoskodtam, hogy a megkezdett mű mielőbb befejezést nyerjen s a nagy feladatok megoldásánál, melyek mezőgazdaságunk fölvirágoztatása érdekében reánk várnak, részletes és megbízható adatok nyujtsanak tájékoztatást s jelöljék ki habozástól és tévedésektől menten intézkedéseink számára a helyes irányt.”*

A motivációk, a célok azóta mit sem változtak. Talán kevésbé bízunk már a mezőgazdaság felvirágoztatásában, de a részletes és megbízható adatokkal továbbra is a gazdaságpolitikai döntések megalapozásához szeretne a statisztika hozzájárulni.

Az első világháború utáni megváltozott helyzet felmérése végett újabb teljes körű összeírásra lett volna szükség, amit a Központi Statisztikai Hivatal 1923-ban, majd 1928-ban is kezdeményezett, de elsősorban finansziális nehézségek miatt csak 1935-ben kerülhetett sor a következő censusra. Ezt a FAO elődjének, a Nemzetközi Mezőgazdasági Intézetnek (International Institute of Agriculture - IIA) a javaslatára hajtották végre. Ebben az összeírásban is részletesen kitértek a mezőgazdaság minden aspektusára (földbirtokok száma, terület, termelési adatok, munkaerő, gépesítettség, gyümölcsfaállomány, állattenyésztés). Az eredmények tükrében nemzetközi összehasonlításokra is sor került. A következő census, szűkebb mutatókörrel, 1942-ben zajlott. A második világháborúban a Hivatal épületeit és berendezéseit közel 40 százalékos kár érte Budapest ostroma alatt. Sajnos az adatgyűjtések és a könyvtár állományának egy része is elveszett. Emiatt az adatok publikálására csak az összeíráshoz képest jóval később, az 1943-1946. évi statisztikai évkönyvben kerülhetett sor. (Oros 1999)

1. ÁBRA: AZ 1895-ÖS CENZUS KIADVÁNYÁNAK CÍMLAPJA

A MAGYAR KORONA ORSZÁGAINAK  
MEZŐGAZDASÁGI STATISZTIKÁJA.

ELSŐ KÖTET.

A MAGYAR MEZŐGAZDASÁGI STATISZTIKA FEJLŐDÉSE  
S AZ 1896. ÉVI VIII. TÖRVÉNYEK ALAPJÁN VÉGREHAJTOTT  
ÖSSZEÍRÁS FŐBB EREDMÉNYEI KÖZSÉGENKINT.



A FÖLDMIVELÉSÜGYI MAGYAR KIR. MINISZTER RENDELETÉBŐL

SZERKESZTI ÉS KIADJA

AZ ORSZÁGOS MAGYAR KIR. STATISZTIKAI HIVATAL.

**Ára 10 forint.**

BUDAPEST  
PESTI KÖNYVNYOMDA-RÉSZVÉNY-TÁRSASÁG.  
1897.

Forrás: <http://kt.lib.pte.hu/cgi-bin/kt.cgi?konyvtar/kt04120602>

A FAO kérése volt, hogy Magyarország is csatlakozzon az 1960-as világcenzushoz. Ez azonban csak részben valósult meg, több lépcsőben, kevesebb területet érintve (1959: országos gyümölcsfa-összeírás; 1960: gépösszeírás; 1965: országos szőlőösszeírás). A FAO igényei egyre hangsúlyosabban érvényesültek, így végül 1970-ben csatlakozott Magyarország először a mezőgazdasági világcenzusok sorához. Az első adatközlés már 7 hónappal az összeírás után megjelent. A census teljes feldolgozásából összesen 15 kötet született, köztük egy angol nyelvű is. Az eredményekből különösen a kistermelésre vonatkozó adatok hoztak újszerű információt.

1980-ban már nem volt kérdés a soron következő világcenzushoz való csatlakozás. A költségtakarékossági szempontok miatt azonban a teljes adatkört az összes nagyüzem megfigyelése mellett, egy teljes körű és 5 reprezentatív felvétellel teljesítették a kistermelői körben. A szintén a FAO adatigények kielégítéséhez végrehajtott 5 évenkénti teljes körű állatszámolásokra 1976-ban és 1986-ban került sor. Ezzel együtt egyéb fontos adatok is begyűjtésre kerültek a földterületre és a mezőgazdasági termelésre vonatkozóan.

A soron következő 1991-es Általános Mezőgazdasági Összeírás is több lépcsőben, több adatgyűjtésből tevődött össze. Az első adatokat tartalmazó kötet 18 hónappal az összeírás után került a nagyközönség elé.

Magyarországon a politikai rendszerváltás a mezőgazdasági termelési viszonyokat érintette a legmélyebben. Gyökeres tulajdonváltás következett be, a termelés súlypontja a korábbi nagyüzemi termelésről a magángazdaságokra helyeződött át. A privatizációs folyamat lényegében befejeződött.

A Központi Statisztikai Hivatal az 1999. évi XLVI. törvény előírásai alapján 2000. március 31-ei eszmei időponttal, 2000. április 1. és 21. között általános mezőgazdasági összeírást hajtott végre. A 2000. évi ÁMÖ a teljes körű mezőgazdasági adatfelvételek sorában a hatodik volt. Az összeírás célja az volt, hogy a magyar mezőgazdaságról pontos és hiteles képet adjon a gazdaságirányítás számára, továbbá információkat szolgáltatson az Európai Unió és Magyarország között folyó csatlakozási tárgyalásokhoz. A felvétel egyúttal kapcsolódott az Egyesült Nemzetek Élelmezési és Mezőgazdasági Világszervezetének 2000. évi világcenzusához is. (Laczka 2000)

A 2000. évi ÁMÖ a mezőgazdasági tevékenységet folytatók teljes körű megfigyelését tűzte ki célul. A végrehajtását elrendelő törvény kimondja, hogy a feladat sikeres teljesítése érdekében adminisztratív nyilvántartások (földhivatali, őstermelői, stb.) gazdálkodót azonosító adatait (név és cím) át kell adni a KSH számára. Tehát már 2000-ben felhasználásra kerültek adminisztratív adatforrások is. Azonban ezek mind minőségi, mind mennyiségi és tartalmi szempontból alacsony szinten álltak még. A teljes körű összeírás során 2,2 millió háztartást (a magyar háztartások közel kétharmadát) kerestek fel a számlalóbiztosok. A felvételek helyszíni munkáit a magyar statisztikai összeírások során már eddig is alkalmazott, jól kipróbált és bevált módszerek alkalmazásával valósította meg a Hivatal. A KSH Mezőgazdasági statisztikai főosztályának irányításával a felvételek végrehajtását az akkori 19

területi (megyei) igazgatóság végezte. A census eredményeinek publikálására tájékoztatási terv készült. Az előzetes adatok közlésére kidolgozott módszer eredményesnek bizonyult, így az adatok közzétételére már az összeírást követő második hónapban sor kerülhetett. A végleges adatok publikálását a KSH 2000 szeptemberétől kezdte meg. Az adatok különféle aggregáltsági szinten, papír formában és elektronikus adathordozón, magyar és angol nyelven is rendelkezésre állnak.

A 2010-es ÁMÖ sorrendben a hetedik census a magyar mezőgazdaság statisztika történetében. A 2010. évi XXIV. törvény a „2010. évi általános mezőgazdasági összeírásról” szabályozza ennek a teljes körű megfigyelésnek a körülményeit, kereteit. A 2010-es census 2010. június 1-jei eszmei időponttal június 1-21. között került végrehajtásra. Az adatok feldolgozása és a publikációk megjelenése még mindig folyamatban van.

### *2.1.2. Szőlő- és gyümölcsös-ültetvények összeírása*

A mezőgazdaság statisztika rendszerének sajátos szeletét adják a szőlő- és gyümölcsös-ültetvények összeírásai. A dolgozat témájához szervesen nem kapcsolódik ez a téma, ugyanakkor a teljesség kedvéért néhány bekezdés erejéig fontosnak tartom, hogy erre is kitérjek.

A IV. századból származik az első, szőlőtermeléssel kapcsolatos irodalmi emlék Magyarország területére vonatkozóan. A telepített szőlőre vonatkozó statisztikai szempontú számbavétel már az 1700-as évek elejétől készül. A legrégebbi szőlőterületet is magába foglaló művelési ágas kimutatás az 1780. évből maradt fenn. A hivatalos statisztikai szolgálat megalakulásától kezdve (1867-től) rendszeres szőlészeti statisztikákat közöltek. Keleti Károly nevéhez fűződik a nemzetközi szőlészeti és borászati statisztika módszertani tervének kidolgozása. A XIX. század végén került végrehajtásra az ültetvény-statisztikai adatgyűjtések tekintetében az első jelentős felmérés, mely a magyar agrárstatisztikának nemzetközi sikert hozott. A legpontosabb, teljes körű eredményeket ezt követően csak az 1960-as években a KSH által végrehajtott, helyszíni felméréseken alapuló összeírás adta. (Ay-Oros 1967) Ezután egészen 2001-ig a mezőgazdasági censusok részeként, illetve a rendszeres összeírásokban kapott helyet a szőlőterület nagyságának és néhány jellemzőjének a megfigyelése. Az összeírás lebonyolítása megteremtette a feltételt az Európai Unió által a tagállamok számára előírt tízévenkénti alapösszeírásokhoz. (Benoist – Falder-Huerta 2002) A 2001-es teljes körű

megfigyelés után 2009-ben zajlott a soron következő nagyobb horderejű adatgyűjtés a szőlő ültetvényekre vonatkozóan, ami tagországgként az első ilyen összeírásunk. A 2009. évi alapösszeírás adatállományának jelentős részét – a Hegyközségek Nemzeti Tanácsa által kezelt szőlőkataszter – külső adatforrás képezi, amely teljes terjedelemben átvételre került. A kataszter, amely a 2001. évi teljes körűen végrehajtott összeírás mutatóira épülve került kialakításra, Magyarország hegyközségi szőlőterületeit tartalmazza és az ország szőlőterületének 80-85 százalékát fedi le. A hazai szabályozás szerint 2008-tól minden 1000 m<sup>2</sup>-t elérő és meghaladó, illetve árutermelő szőlőültetvény része a hegyközségi nyilvántartásnak. Ennél kisebb vagy nem kereskedelmi céllal létesített szőlőültetvények használóinak nincs bejelentési kötelezettségük a hegyközségek felé. A hegyközségi területeken kívüli szőlőültetvények (a 2001. évi felmérés szerint 10 231 hektár és közel 89 ezer ültetvény) reprezentatív mintán, összeírói hálózat segítségével kerültek megfigyelésre. A minta alapján történő megfigyelés keretében közel 1 000 hektáros területen megközelítően 10 ezer ültetvény összeírására került sor 2009. június 1. és szeptember 30. közötti időszakban.

A gyümölcsösök statisztikai összeírása még a szőlőültetvényeknél is nehezebb feladatot jelent. Sok a konyhakertekben, út mentén, szőlősorok között termő gyümölcsfa, amelyeknek mind a területét, mind a termését nehéz számba venni. A hivatalos statisztika csak viszonylag későn, az 1895-ös általános mezőgazdasági összeírás alkalmával kérdezte meg először a gyümölcsfák számát, csakúgy, mint az 1935. évben. Önálló művelési ággként csak az 1950-es évektől szerepelnek a gyümölcsösök, a második világháborút követő nagyobb volumenű ültetéseknek köszönhetően. 1959-ben zajlott az első gyümölcsfa-összeírás a KSH keretei között. Ezután a nagy horderejű összeírás után az 1972-es censusig csak az adatok továbbvezetése, korrekciója zajlott, illetve az árutermelő gyümölcsösök legfőbb adatait gyűjtötték 1963-tól folyamatosan. A 2001-es szőlő- és gyümölcsösültetvény-összeírás sok tekintetben különleges munkának számít a KSH mezőgazdasági felvételei között. Egyrészt a két-fázisos munka miatt: első fázisban a tulajdonosokat és adataikat írták össze a számlálóbiztosok, második fázisban pedig az ültetvények helyszíni összeírására került sor képzett szakembergárda segítségével. (A két fázis az adatok rögzítését és adatbázis szinten az összekapcsolását is nagyon megnehezítették. Magyarországon történeti, szabályozási okokból olyan sajátos helyzet alakult ki, hogy egy-egy ültetvénynek több, esetenként akár 10-20 tulajdonosa is van. Az ültetvény viszont ténylegesen nincsen felosztva. Így sok gondot okozott a többszörös tulajdonlás kezelése.) Másrészt az összeíráshoz felhasznált segédanyagok miatt is speciális volt ez az adatgyűjtés. Az ÁMÖ-ből származó, lajstromba

szervezett tulajdonosi adatok kiegészültek több adminisztratív forrás információival. Emellett pedig a Földmérési és Távérzékelési Intézet jóvoltából ún. felt- és kataszteri térképeket is felhasználtak az ültetvények helyszíni felmérése során. (Laczka 2002)

A 2001. évi – valamennyi gyümölcsfajra kiterjedő – teljes körű felmérést követően 2007-ben négy gyümölcsfajra (alma, körte, őszibarack és kajsziarack) vonatkozó összeírást hajtott végre a Hivatal. A korábbi census adataira és az ültetvénytelepítési információkra épülve reprezentatív megfigyelés keretében valósult meg ez a felmérés. A mintavételi terv két részből tevődött össze. Egyrészt a 2001-es adatok alapján az ültetvényeket méretük szerint csoportosították. A viszonylag nagy ültetvényekkel rendelkező települések közel 100 százalékban bekerültek a mintába. Az ott található kisebb ültetvények reprezentálták az alacsonyabb méretkategóriát. Másrészt – a Mezőgazdasági és Vidékfejlesztési Hivataltól kapott – a támogatás igénybevételével tervezett ültetvénytelepítésekre vonatkozó információk alapján egy úgynevezett kiegészítő mintát is lekérdeztek.

Az ültetvények számbavétele tehát egy speciális része a mezőgazdasági összeírásoknak. Egyrészt az ültetvények állandósága, és jellege speciális, másrészt sokáig csak teljes körű összeírásokkal történtek adatgyűjtések. Mintavételes megfigyelésekre csak az Európai Unió jogszabályok alapján, először gyümölcsösökre a 2007-es, szőlőre a 2009-es adatgyűjtésben került sor. A mintavétel módszere is merőben eltér a termelésstatisztika többi területétől a jellegzetességek miatt.

### *2.1.3. Állatállomány összeírások*

Az állattenyésztés fontosságának megfelelően már nagyon korán megfogalmazódott az ágazatra vonatkozó adatigény. Már a törökök által megszállt területeken, az adózás alapját képező defterek is kiterjedtek az állatállományra. Csak úgy, mint a XVIII. század végén, a XIX. század elején végrehajtott adóösszeírások, valamint az osztrák adminisztráció statisztikai célú megfigyelései. Magyar keretek között a statisztikai adatok gyűjtése több mint 150 éves múltra tekint vissza. Eleinte a népszámlálásokkal egyidejűleg kerültek lebonyolításra az állatszámítások. 1884-től önálló összeírásként szerepelt, és természetesen a rendszeresen ismétlődő mezőgazdasági censusok kérdőívének is részét képezte egy állatállománnyal foglalkozó blokk.

A 2. ábra egy XIX. század végi szarvasmarha állomány lekérdezésére kialakított táblázatot mutat be. Ebben – a kor berendezkedésének megfelelően – a tulajdonosi forma szerint is megbontották az állomány nagyságát, valamint rákérdeztek az állatok értékére is. A 3. ábraán láthatunk egy aktuális táblázatot napjaink kérdőíveiből. A kor és ivar szerinti bontás valamivel részletesebb, viszont a fajtákra nem kérdezik rá. A pénzbeli adatok pedig igencsak kényes témának számítanak, ezért csak az állomány nagysága képezi a kérdések tárgyát.

Természetesen a kérdésfeltevés módja sokat változott. A tulajdonosi viszonyok már nem jelentenek csoportosítási szempontot. A kor és ivar szerinti bontás pedig tovább finomodott a részletesebb adatigényeknek megfelelően. Azt is láthatjuk, hogy a „modern” kérdőívek már valamilyen nomenklatúrán alapulnak, és minden tételnek saját kódja van. A XIX. század végén ez még nem volt jellemző.

2. ÁBRA: AZ 1895-ÖS CENZUS KÉRDŐÍVÉNEK EGYIK ÁLLATÁLLOMÁNNYAL FOGLALKOZÓ TÁBLÁJA

**I. Szarvasmarha.**  
**1. Magyar fajta**  
(hosszú- és rövidszarvú és riska).

		Tenyész- bika	Tehén	Ökör	Növendék				Összeg
					1 éven aluli	2 éven aluli	3 éven aluli	4 éven aluli	
Van darab	paraszt-birtokos tulajdonában								
	közép- és nagybirto- kos tulajdonában								
Átlag- érték	paraszt-birtokos tulajdonában								
	közép- és nagybirto- kos tulajdonában								

Forrás: <http://kt.lib.pte.hu/cgi-bin/kt.cgi?konyvtar/kt04120602>

Az 1895-ös általános mezőgazdasági összeírás után 1911-ben hajtották végre a következő állatszámlálást. A Statisztikai Hivatal ezután 1935-ben és 1942-ben végezte el az állatállomány teljes körű összeírását. 1945 és 1947 között a háborús károk mértékének megállapítása céljából az állatállományt községenként évente többször is becsülték. Az összeírások közötti időszakokban a Földművelésügyi Minisztérium megfigyelései alapján tájékozódtak.



Az 1930-as évek végétől voltak ún. reprezentatív sertésösszeírások az akkori Gazdaságkutató Intézet szervezésében, ahol előrejelzések készítéséhez használták fel az adatokat. A II. világháborúig a háziállatokat minden tavasszal összeírták az állatok legelőre hajtásakor a közigazgatási hatóságok. Ezen adatok minősége messze elmarad a jelenlegi igényektől, hiszen nem gazdálkodónként történt a számbavétel, és a kihajtott állatok állományát együttesen, sokszor csak becslésre alapozva jegyezték fel. Az adatgyűjtések maihoz hasonló rendszere az 1940-es évek végén alakult ki. 1949-től a tavaszi teljes körű állatszámolás mellett mintavételes összeírásokra is sor került. (Oros 1954) Ezeket a KSH hajtotta végre. Előbb évi kettő, később évi három adatgyűjtés volt. 1950-1973 között minden év tavaszán teljes körű állatszámolást tartottak. 1974-től tértek át a decemberi referencia időpontra, és ritkulni kezdtek a censzusok is. (Laczka et. al. 1998)

### 3. ÁBRA: A 2007-ES GSZŐ EGYIK TÁBLÁJA

#### 101. Szarvasmarha

101. Szarvasmarha				
Kód	Megnevezés			Darab
2014	1 évesnél fiatalabb borjú	vágásra	hímivarú	
2015			nőivarú	
2041		egyéb tartásra	hímivarú	
2021			nőivarú	
2042	1–2 éves szarvasmarha	hímivarú		
2103		nőivarú	vágásra (vágóüsző)	
2024			egyéb	
2043	2 évesnél idősebb szarvasmarha	hímivarú		
2104		üsző	vágásra (vágóüsző)	
2026			egyéb (előhasi üsző)	
2031		tehén	tejhasznú	
2034			húshasznú	
2033			kettős hasznosítású	
2000	Szarvasmarha összesen			

*Forrás: 1677/07 OSAP számú kérdőív*

A szocialista mezőgazdasági nagyüzemek kialakulása, illetve ezek rendszeres statisztikai beszámoltatása miatt 1957-től már csak az egyéni gazdaságok állományát figyelik meg reprezentatív megfigyelésekkel. A nagyüzemek, gazdálkodó szervezetek adatai eleinte minden negyedév végén, majd évente háromszor postai úton érkeznek be. A mintavételes összeírásokat a költségtakarékosság tette szükségessé, ugyanakkor az egyre komolyabb statisztikai módszerek alkalmazása tette lehetővé. (Ay 1976)

A mintavételi eljárások folyamatosan fejlődtek. A módszerek a különböző időszakokban a következőkben bemutatottak szerint alakultak (Schindele 1989):



- **1949-1959:** Teljes községeket választottak ki a mintába. A minta kijelölése területi módszerrel történt: térképre fektetett négyzetháló metszéspontjai jelölték ki az elsődleges megfigyelési egységeket, majd a településeket nagyságuk szerint rétegezték. A mintavételi arány 30 százalékos volt. A módszer előnye volt az egyenletes földrajzi lefedettség, ugyanakkor az országon belüli eltérő településszerkezet hátrányosan befolyásolta a kiválasztást. A minta így az intenzívebb sertéstartással jellemezhető Alföldön arányaiban nagyobb volt, a becslés felfelé torzított volt.
- **1960-1969:** A sertéstenyésztés jellege, pontosabban a kocák aránya határozta meg a kiválasztást, hiszen ez jó indikátora a teljes éves sertésállománynak. Továbbra is teljes községek kerültek a mintába, és megmaradt a mintavételi arány. A kiválasztott településeket megyéenként a kocaállomány aránya szerint növekvő sorrendbe rakták, így alakítottak ki három nagyságkategóriát, amelyekből ezután arányosan kerültek be a települések.
- **1970-1975:** A censusok során kialakított úgynevezett összeírási vagy számláló körzeteken alapuló kiválasztásra tértek át. Egy körzet területileg viszonylag összefüggő volt, 150-200 állattartó háztartással. A körzetek kijelölése a korábbi kocaállomány szerinti nagyságkategóriás rétegzéssel, arányosan történt. A mintavételi arány 10 százalékra csökkent. Viszont jelentősen szétszórtabb mintasokaság állt elő, így összességében az összeírási költségek jelentősen nem csökkentek.
- **1976-1995:** A mintakörzeteket a sertéstartás intenzitása szerint csoportosították. A körzetekből két – 12 címet tartalmazó – csoport került véletlen módszerrel kiválasztásra. A minta alapsokasághoz viszonyított aránya tovább csökkent, 4,5 százalékra.
- **1995-től** a maihoz hasonlóan a gazdaság nagysága alapján történt a rétegzés, továbbra is körzet alapú kiválasztást követően.

Az adatok felszorozása hányadosbecsléssel, illetve regressziós becsléssel történt. A becslések relatív standard hibája 1 százalék körül mozgott, vagyis meglehetősen jó minőségű adatokat szolgáltatottak a reprezentatív megfigyelések. Az országos számokon kívül megyei szintre is elkészítették a becsléseket.

Az 1990-es években volt egy más jellegű vonulata is a mintavételi módszerek alkalmazásának (Szemán et. al. 1998). Az egyéni gazdaságokat 7 nagyságkategóriába sorolták a termőterület és a számosállatban mért állatállomány nagysága szerint. Ezen belül az állatállomány megfigyeléseket fajok szerint külön kérdőíven és külön almintán hajtották végre. Ezeket a próbálkozásokat azonban több kritika érte, ezért csak rövid ideig használták ezt a módszert, és a későbbiekben sem volt példa hasonló irányok bevezetésére.

Az összeírások már az 1960-as évektől minden tenyészállatfajra kiterjedtek, 1986-tól kor és ivar szerinti bontásban is. 1991-ig 5 évente hajtottak végre tételes, teljes körű számbavételt az ország teljes állatállományáról. A közbeeső időszakokban pedig negyedévenként reprezentatív mintavételt alkalmaztak a megfigyelésekhez. (Nyitrai 1996) Az éven belüli többszöri megfigyeléseket az állatállomány-alakulás szezonális jellege tette indokolttá. Ezen belül is elsősorban a kistermelés legnagyobb részét adó sertésállomány ciklikus ingadozása. A sertésállomány közel felét ugyanis a kistermelők adják, ők pedig viszonylag gyorsan képesek reagálni a piaci változásokra. „A nagy sertésállomány miatti túlkínálat lenyomja a sertésárakat, ennek hatására a tenyésztők csökkentik kocáik számát, ami a sertésállomány visszaesését eredményezi. Az alacsony állatállomány miatt nő a kereslet, magasra emelkedik a sertés piaci ára, ez visszaállítja a tenyésztési kedvet, a gazdák ismét növelik a kocák számát, amit a sertésállomány felfutása követ. A megnövekedett állomány ismét lenyomja a sertésárakat és a ciklus mozgás tovább folytatódik. A fenti folyamat a sertésciklus, amely az 1970-es évek közepéig szabályosan ismétlődött. Attól kezdve az agrárkormányzat, a negyedéves adatok ismeretében hathatós intézkedéseket tett a sertésciklus kivédésére.” (Schindele 1989. 364. p.) Az 1980-as évek végétől a sertésciklus kevésbé jelentkezett. A kormányzati intézkedések mellett ebben az is közrejátszott, hogy egy sertés átlagos hízalási ideje mintegy felére csökkent le. A ciklus 3 évenkénti szabályos impulzusait ez lényegesen lerövidítette. Éppen ezért ritkítani lehetett a megfigyelések gyakoriságát. Így az évi 3 időpontra módosult. 2009-től áttért a Hivatal az évi 2 összeírára (az EU-s előírásoknak megfelelően). A december 1-jei eszmei időpontra vonatkozó állatállomány adatok azonban így is minden időszakból rendelkezésre állnak, így – bár az éven belüli ingadozás monitorozása nehezebb – az idősorok folytonossága továbbra is biztosított.

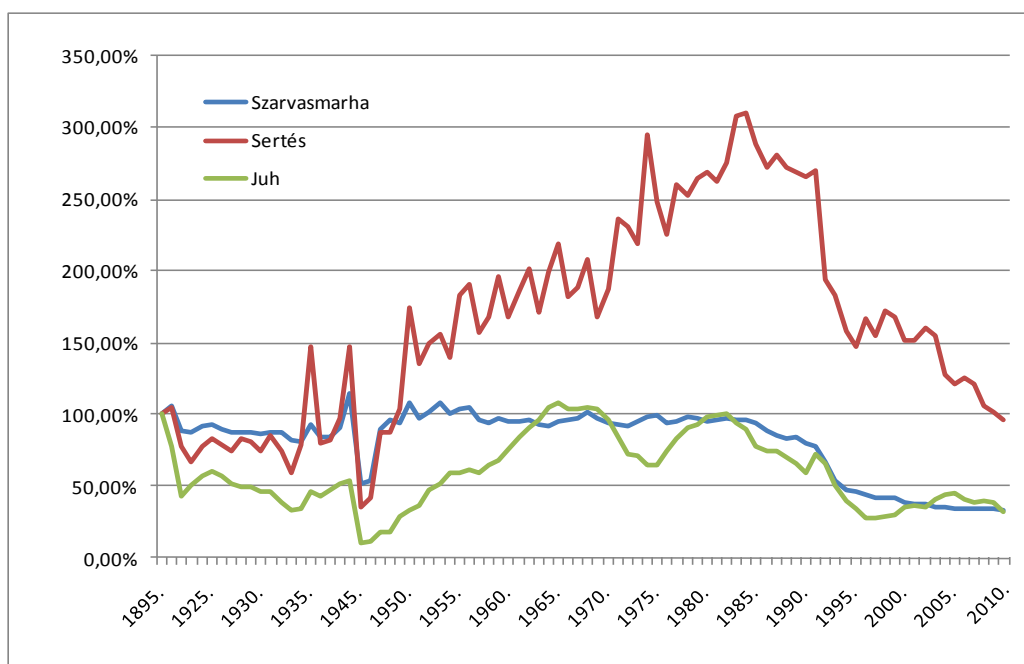
1997-ben (bár még nem voltunk tagok) került sor az első olyan állatállomány összeírára, ami már megfelelt az Európai Unió követelményeinek is. Azóta folyamatosan figyelembe

veszi és teljesíti az EU-s jogszabályok előírásait a Hivatal. Többek között a módosuló előírások miatt kezdtem el jelen kutatás során is az állatállománnyal foglalkozni.

## 2.2. A magyar mezőgazdaság teljesítményének időbeli alakulása

A mezőgazdasági statisztika jelentős múltjának, a rendszeres censzusoknak és évközi reprezentatív megfigyeléseknek köszönhetően több mint 100 éves hosszúságú idősorokat találunk a KSH honlapján. A hosszú távú változások áttekintéséhez gyűjtöttem össze a dolgozat szempontjából fontos állatfajok állományának alakulását, és néhány művelési ág területének adatait. (A részletes adatokat az I. és II. számú melléklet tartalmazza.) A jellemző tendenciák elemzésének alapját a hosszú idősorok adatai, valamint Buday-Sántha (2011) adják.

4. ÁBRA: 3 ÁLLATFAJ ÁLLOMÁNYÁNAK ALAKULÁSA MAGYARORSZÁGON, 1895-2010  
(1895=100%)



Forrás: KSH honlap: „Agrár idősorok és censzusok” alapján saját számítás

Az állattenyésztéssel foglalkozó ágazatok szerepének jelentősége a megtermelt növényi termékek értékesebb állati termékekké való átalakításában rejlik. Emellett a mezőgazdasági melléktermékeket is hasznosítják. Az elmúlt 115 évben az állatállomány jelentős

változásokon ment keresztül. Az 4. ábra bázisviszonyszámokat ábrázol a közös nevező, az áttekinthetőség és a könnyebb összehasonlíthatóság céljából. A háborús éveket leszámítva a szarvasmarhák száma mutatja a legnagyobb stabilitást. A leghektikusabb mozgás a juhállományánál mutatkozik, ami képes nagymértékű változásokra akár néhány év leforgása alatt is. Ez az állomány pontos becslésében napjainkig komoly problémát okoz. Mindhárom állatfajra igaz, hogy a rendszerváltás környékétől kezdve szinte folyamatos és radikális csökkenésen mentek keresztül. 2010-re az XIX. század végi állatszámokat még a juhok sem érik el, a sertések és a szarvasmarhák pedig még annak a felét sem közelítik meg.

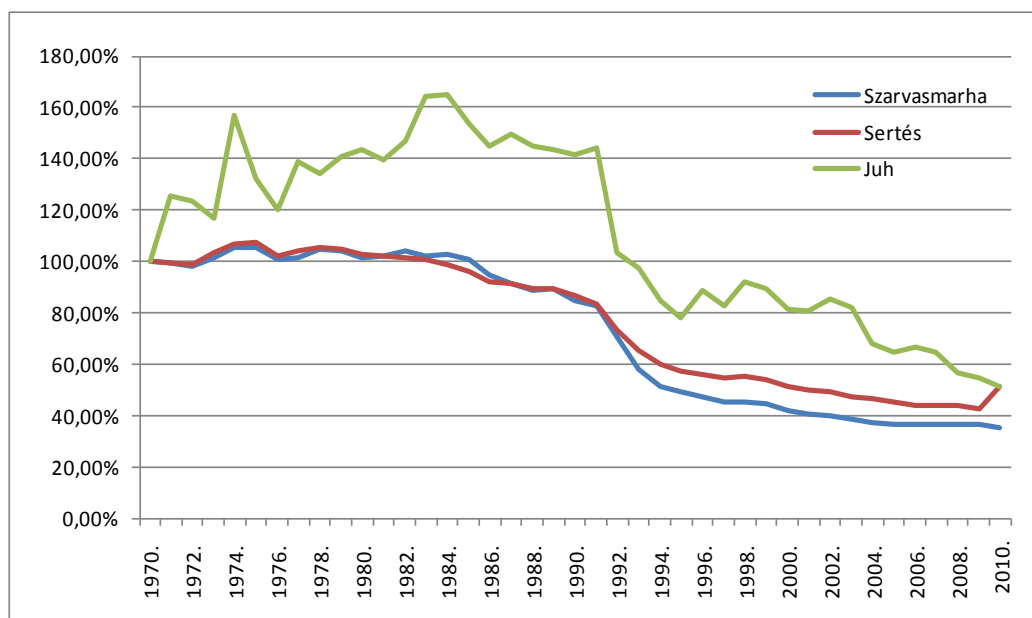
A közelmúlt folyamatainak áttekintéséhez 1970-es bázisra is átszámítottam az értékeket. (Lásd 5. ábra.) Ebből az látható, hogy még a 70-es évek szintjéhez képest is mindhárom állatfaj állománya a felére, vagy még annál is alacsonyabb szintre esett vissza. A sertésállomány 2005-ben, közel 60 év után esett 4 millió alá. Az állattartás egyre nehezebb gazdasági feltételei és körülményei egyértelműen kedvezőtlenül hatnak a hazai termelésű állatok mennyiségének alakulására. A hivatalos statisztika szempontjából ugyanakkor fontos észrevennünk azt a tényt is, hogy a szarvasmarha- és a sertésállomány a juhokhoz képest viszonylag kiegyensúlyozott időbeli alakulást – csökkenést – mutat az elmúlt 10 évben. Ez a gazdaságok egyfajta stabilitását sugallhatja, bár strukturális változások azért meghúzódnak a háttérben. A stabilitás az összeírások megtervezése, lebonyolítása és hatékonysága szempontjából egy nagyon fontos tényező, ami a hivatalos statisztikai adatok minőségére jótékony hatással lehet. Az állattenyésztést érintette leginkább a magyar mezőgazdaság elhúzódó válsága. Az állományok csökkenése még nem állt meg, de lelassulni látszik a fogyás üteme.

Az ország éghajlata, fekvése jó termőképességű talajok kialakulását tette lehetővé. Az ország területének 83 százaléka a mai termőterület aránya. A művelésből kivont területek ugyanakkor folyamatosan bővülnek. Ez nem magyar sajátosság, hanem világtendencia. Oka a fejlődés, és a vele járó település-, ipar- és infrastruktúra fejlesztés területigénye.

A földterület alakulása művelési ágak szerint természetesen az állatállományhoz képest jóval „simább” képet mutat. (Lásd 6. ábra.) A földterület sokkal rugalmatlanabb, már ami a művelési ág váltást illeti. Legfeljebb a szántóföldi növények struktúráját (a vetésszerkezetet) lehet gyorsabb ütemben változtatni. Az állandó ültetvények, mint a szőlő és a gyümölcsös pedig természetüknél fogva hosszabb időn át változatlanok. Éppen ezért, statisztikai, mintavételi szempontból nem is a művelési ágak területének becslése támasztja a

nehézségeket, hanem azok vetésszerkezetének, a konkrét növények termőterületének és betakarított mennyiségének a becslése.

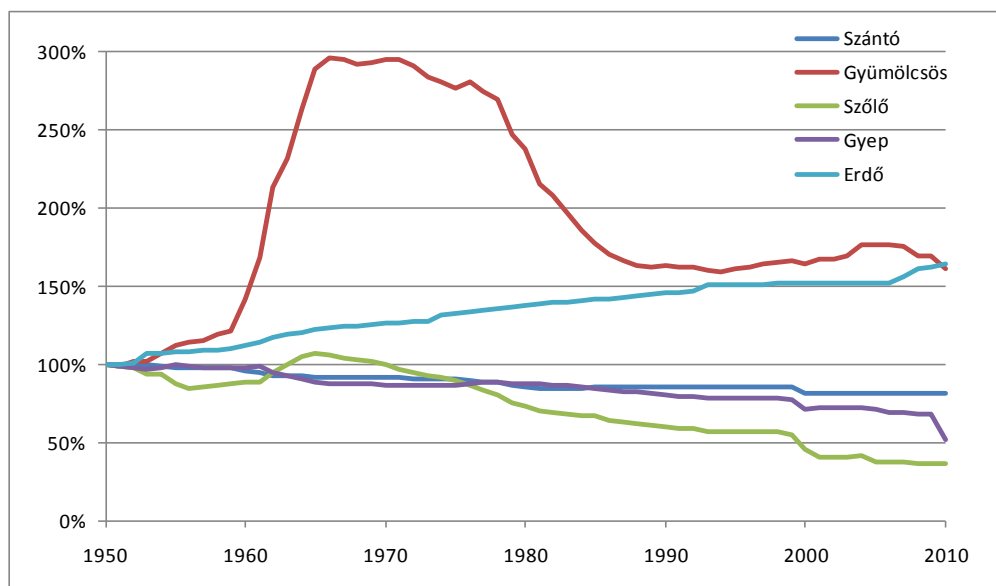
5. ÁBRA: 3 ÁLLATFAJ ÁLLOMÁNYÁNAK ALAKULÁSA MAGYARORSZÁGON, 1970-2010  
(DECEMBER 1.) (1970=100%)



Forrás: KSH honlap: „Agrár idősorok és censzusok” alapján saját számítás

A gyümölcsös a '60-as, '70-es évek szárnyalása után az elmúlt 25 évben 100 ezer hektár körül, az 1950-es állapot 1,7-szeresén állapodott meg. Ez az egyik legintenzívebb művelési ág. Magyarországon a jelenlegi terület körülbelül 4-szerese lenne alkalmas erre a célra. Az alacsony szintű tőke és a megfelelő termelési rendszerek hiánya azonban nem teszi lehetővé ennek a kihasználását. A szőlőterület folyamatos zsugorodás után 80-90 ezer hektár között stabilizálódott az elmúlt 10 évben. A gyümölcsöshöz hasonlóan, az alkalmas terület nagysága ennél jóval magasabb. Ugyanakkor a terület csökkenését a korszerűbb ültetvények hozamaival ellensúlyozni lehet. A gyepek (legelő és rét) és a szántóterület nagyjából hasonló mértékű csökkenésen ment keresztül. Ezzel párhuzamos, ha lassan is, de fokozatosan csökken az összes területen belül elfoglalt arányuk. A gyepek napjainkban kezd erőteljesebben visszaesni. Az egyetlen, viszonylag dinamikusan növekvő művelési ág az erdő. Területe mára megközelíti a 2 millió hektárt. 60 év alatt, több mint 1,5-szeresére gyarapodott. Ennek egyik oka, hogy az erdő művelési ág jelentősége gazdasági, társadalmi és ökológiai szempontból is egyre inkább növekszik.

6. ÁBRA: A FÖLDTERÜLET ALAKULÁSA FŐBB MŰVELÉSI ÁGAK SZERINT, MAGYARORSZÁGON,  
1950-2010 (1950=100%)



*Forrás: KSH honlap: „Agrár idősorok és censzusok” alapján saját számítás*

### 2.3. A statisztika jogi hátterének alakulása

Az összeírások történetéhez az is hozzátartozik, hogy az idők során a statisztikai munka jogszabályi háttere hogyan alakult. Úgy gondolom nem haszontalan ezért egy pár oldal erejéig erről is beszélni.

Magyarország történetében eddig 6 statisztikai törvény született:

- 1874. évi XXV. törvény-cikk az országos statistika ügyének szervezéséről
- 1897. évi XXXV. törvény-cikk a M. Kir. Központi Statisztikai hivatalról
- 1929. évi XIX. törvény-cikk a hivatalos statisztikai szolgálatról
- 1952. évi VI. törvény az állami statisztikáról
- 1973. évi V. törvény a statisztikáról
- 1993. évi XLVI. törvény a statisztikáról

Az első hazai statisztikai törvény 1874-ben jelent meg, és ezzel Magyarország a statisztika jogi szabályozásának területén az európai élvonalon álló országok közé tartozott.

A nagy jelentőségű első statisztikai törvény után, 1897-ben fogadták el a XXXV. törvénycikket, a második statisztikai törvényt, aminek máig tartó hatása van, és amit részletes hivatali ügykör és ügyviteli szabályzat egészített ki. Itt fogalmazódott meg a Hivatal új megnevezése: Magyar Királyi Központi Statisztikai Hivatal. Részletesen szabályozásra került az igazgatójának, és a fölé rendelt kereskedelemügyi miniszternek a hatásköre is.

Mai szemmel érdekes látni, hogy az 1897-es törvény rendelkezik az összeírók személyéről is. Előírja ugyanis, hogy az iskolák férfi tanítói kötelesek összeíróként közreműködni az adatgyűjtéseknél a kereskedelemügyi miniszter rendelete alapján. Ennek megfelelően ezeket az összeírásokat célszerű az iskolai szünidőkkel összehangoltan szervezni, illetve 6 nap egy évben lehet tanítási szünet ilyen indokkal. Ugyanezt találjuk az 1929-es törvényben is.

Az 1929. évi XIX. törvénycikk a hivatalos statisztikai szolgálatról sorrendben a harmadik statisztikai törvény. Ez már kiterjed a statisztikai szolgálatra, aminek elsődleges célja az volt, hogy a párhuzamos adatgyűjtéseket visszaszorítsák. Emellett megváltoztak a KSH-t felügyelő szervek. A Hivatal közvetlenül a miniszterelnök felügyelete alá került. Ez olyan aggályokra adott okot, mint a szakmai szempontok háttérbe szorulása, illetve a politikai szempontok érvényesítésének veszélye a statisztikai munkában.

Az 1952-es év nagy jelentőségű a magyar statisztikai szolgálat életében. Ekkor ugyanis az évtizedek óta tartó törekvések eredményeképpen megalakulhattak a KSH területi szervei, a megyei igazgatóságok és a járási, városi felügyelőségek. (A Fővárosi Statisztikai Hivatal már korán megalakult, de a főváros területére korlátozódott a működése és csak 1950-től rendelték a Hivatal alá.) Ennek legnagyobb jelentősége az adatfelvételekhez nélkülözhetetlen helyi ismeretekben rejtett, és nem utolsó szempont volt az sem, hogy gyakran a területi munkatársakból kerültek ki a helyi és központi vezetők.

Az '52-es és a '73-as törvény a szocializmus égisze alatt született. Ennek megfelelően alakult a tartalmuk is. Az előzőekhez képest lényegesen eltérő képet mutatnak ezek a jogszabályok. A társadalmi, politikai és gazdasági helyzet megváltozásával a statisztika szerepe is átalakult. Az 1952-es törvény megfogalmazásában: „*A statisztika a Magyar Népköztársaságban a gazdasági, szociális és kulturális fejlődés ellenőrzésnek és tervszerű irányításának eszköze.*” Az 1973-as törvény ezt ugyan árnyaltabban fejezi ki, de mindkét helyen megjelenik a KSH ellenőrző szerepe a statisztikai adatokat illetően, amire se korábban, se később nem találunk példát.

Az 1993-as törvény legnagyobb részt az 1929-es statisztikai törvényből táplálkozik. Mivel az adatvédelmi törvény egyik alaptétele az adatgyűjtések célhoz kötöttsége, ezért a statisztikai törvény elsődleges feladata kell, hogy legyen a statisztikai cél fogalmának tisztázása. Egy ilyen definíció pontos leírása rövid, tömör formában meglehetősen nehéz feladat. Nyilvánvalóan cél a társadalom és a gazdaság állapotának, változásának bemutatása. Emellett azonban az is fontos, hogy az adatkezelés statisztikai céljáról akkor beszélünk, ha nem az adatok egyedisége számít, hanem több adat egybefoglalt értéke. A törvény 1. paragrafusa tisztázza tehát a cél fogalmát, és felsorolja a statisztikai tevékenységbe tartozó munkafolyamatokat: adatok felvétele, feldolgozása, tárolása, átadása, átvétele, elemzése, szolgáltatása, közlése és közzététele.

A 3. paragrafusban a jogszabály alkotók felsorolják, hogy mely szervek tartoznak a hivatalos statisztikai szolgálathoz. Ezen szervek köre túllép a végrehajtó hatalom szervein, hiszen ide tartozik többek között a Legfelsőbb Bíróság és a Magyar Nemzeti Bank is. A Központi Statisztikai Hivatal tehát nincs egyedül az információszerzés rendszerében, bár kitüntetett szakmai szerepe van. A törvény által felsorolt szervekre is ugyanúgy vonatkoznak a statisztikai céllal és az adatok kezelésével kapcsolatos szabályok, sőt a KSH állásfoglalását is figyelembe kell venniük.<sup>4</sup> Ezt, a több intézményi részvételt látjuk a mezőgazdasági adatgyűjtések terén is.

Az Országos Statisztikai Tanács (OST) tagjai a hivatalos statisztikai szolgálathoz tartozó szervek, az érdekképviselők, az önkormányzatok, a társadalombiztosítás és a tudományos élet képviselőiből álló testület. Feladatát a törvény 7. paragrafusa tisztázza: társadalmi érdekek és adatszolgáltatói igények képviselője, szakmai tanácsadás a KSH elnökének, valamint az Országos Statisztikai Adatgyűjtő Program véleményezése. Ülésein az adatvédelmi biztos állandó meghívottként vesz részt, mintegy kihangsúlyozva az adatvédelemmel kapcsolatos kérdések jelentőségét.

A statisztikai munka részletes szabályozását kormányrendeletekre bízta a törvény. Ez pedig nem más, mint az Országos Statisztikai Adatgyűjtési Program (OSAP), ami évente, részletesen és egységesen szabályozza az adatgyűjtések körét.

---

<sup>4</sup> 170/1993. (XII.3.) Korm. Rendelet a statisztikáról szóló 1993. évi XLVI. törvény végrehajtásáról



A gyakorlati lebonyolítás egységességének biztosítása érdekében a törvénynek megfelelően a Vhr. 6. paragrafusa tartalmazza a kérdőívekre vonatkozó kötelező formai kellékeket, követelményeket is.

A publikációs tevékenység szabályozása a 22-24. paragrafusokban található. Ennek lényege, hogy egy összehangolt tájékoztatási rendszert kell kialakítani, amelyhez az OST koordináló tevékenysége nyújt segítséget. A KSH kiemelt szerepe itt is megjelenik, hiszen a nemzetközi szervezeteknek történő tájokoztatás szervezése a KSH felelősségi körébe tartozik.

Az Eurostat az Európai Unió statisztikai hivatala. Munkájának kereteit a Tanács 1997. február 17-i 322/97/EK rendelete a „közösségi statisztikákról”, a Bizottság 1997. április 21-i 97/281/EK határozata „az Eurostat szerepéről a közösségi statisztikák elkészítésében” jogszabályi háttér biztosítja. Szervezetileg az Európai Bizottság egyik főigazgatósága. Luxemburgban székel.

A tagországok statisztikai intézményeivel kell folyamatos kapcsolatban lennie. Egyrészt a módszertani egyeztetések miatt, másrészt tőlük kell begyűjtenie az adatokat, amelyek szolgáltatására jogszabályi előírások vonatkoznak, vagy úgynevezett „gentlemen’s agreement” alapján történnek. (Nagyon érdekes, hogy az adatbegyűjtések közel 40%-a alapul informális megállapodásokon a törvényhozás bonyolultsága és időigényessége miatt. Ráadásul viszonylag jól is működnek az együttműködések még ilyen alapon is.)

Az Európai Unióban az Európai Statisztikai Rendszer (European Statistical System – ESS) biztosítja a statisztikai munka alapját. Ennek sok résztvevője van: a tagállamok statisztikai hivatalai, a nemzeti központi bankok, az Európai Központi Bank (European Central Bank), a Statisztikai Program Bizottság (Statistical Programme Committee), a Monetáris, Pénzügyi és Költségvetési Statisztikai Bizottság (Committee on Monetary, Financial and Balance of Payment Statistics), a különböző munkacsoportok, döntés-előkészítő testületek, és természetesen az Eurostat. Ők közösen alakítják ki hosszú egyeztetési folyamaton keresztül a

Többéves Statisztikai Programot (Multiannual Statistical Programme)<sup>5</sup>, ami 5 évet fog át. Ez kerül lebontásra Éves Statisztikai Programmá az operatív munkához.

A KSH feladatait nagyon sokban érintik az Eurostat elvárásai és a harmonizáció teljesítése. A sok éves gyakorlatot számos területen újra kellett gondolni az elmúlt pár évben az egységes előírásoknak való megfelelés érdekében.

A statisztikai munkát és az adatgyűjtés, publikálás folyamatát még egy fontos jogszabály keretezi, az egyik alkotmányos emberi jog védelmében. Az Alkotmány 59. paragrafusa kimondja, hogy mindenkit megillet a személyes adatok védelméhez való jog. A másik oldalról megközelítve (61§ (1)) viszont mindenkinek joga van megismerni a közérdekű adatokat. A magyar törvényalkotásban az adatvédelem és az információszabadság szabályozását egy jogszabály keretein belül foglalták össze. Ennek megfelelően egy általános, keret jellegű törvény – az 1992. évi LXIII. törvény „a személyes adatok védelméről és a közérdekű adatok nyilvánosságáról” – született.

Az adatok továbbítására és összekapcsolására is szigorú szabályok vonatkoznak. Az érintettek hozzájárulása vagy törvényi felhatalmazás szükséges hozzá. Mivel az adatok továbbítása, illetve összekapcsolása is része az adatkezelésnek, ezért ezekben a szakaszokban is meg kell felelni az előre definiált céloknak. Még szervezetén belül is korlátozza az adatok összekapcsolását az adatvédelmi törvény. A statisztikai rendszer működése és sajátosságai azonban megkövetelik, hogy a rendelkezésre álló adatvagyonról a költséghatékonyságnak és az ésszerűségnek megfelelően használjuk ki. Mivel a statisztika jellegénél fogva a feldolgozások során nem az egyedi adatok az érdekesek, hanem azok bizonyos szintű aggregátumai, ezért speciális szabályok születtek a statisztikai célú adatfelhasználás tekintetében. A háromnál kevesebb adatszolgáltatóra vonatkozó adatokat nem szabad publikálni, és természetesen egyedi adatok közzétételére sincsen jogszabályi felhatalmazás. De nem is elsősorban a publikálás terén jelent megszorítást az adatok védelme, sokkal inkább a begyűjtés oldalán. A kérdésfeltevések módját, a kérhető adatok körét is befolyásolják az adatvédelmi előírások.

---

<sup>5</sup> Decision No 2367/2002/EC of the European Parliament and of the Council of 16 December 2002 on the Community statistical programme 2003 to 2007

A statisztikai munka alapvető kereteit, háttérét tehát számos jogszabály meghatározza és szabályozza. Ebbe a keretrendszerbe ágyazva születnek meg a szakterületekre vonatkozó előírások, és persze a gyakorlati lebonyolítás rendszere.

Az általános jogszabályi háttér áttekintése után a következő fejezetben áttérek arra, hogy mindez mit jelent a mezőgazdaság statisztika számára.

## **2.4. A mezőgazdasági termelésstatisztika rendszere**

### **2.4.1. A megfigyelési egységek**

A 1166/2008/EK rendelet<sup>6</sup> definiálja a mezőgazdasági gazdaság fogalmát (az 1988-as definíció alapján):

„Mezőgazdasági üzem vagy „gazdaság”: önálló irányítású, technikailag és gazdaságilag különálló egység, amely az I. mellékletben felsorolt mezőgazdasági tevékenységeket végez az Európai Unió gazdasági területén, akár főtevékenységként, akár melléktevékenységként.”

A jogszabály I. melléklete szerint a NACE 2. módosítása szerinti következő tevékenységek tartoznak ide:

- 01.1: Nem évelő növények termesztése
- 01.2: Évelő növények termesztése
- 01.3: Növényi szaporítóanyag termesztése
- 01.4: Állattenyésztés
- 01.5: Vegyes gazdálkodás
- 01.6: Mezőgazdasági, betakarítást követő szolgáltatás.

A KSH gyakorlata szerint a fenti definíciónak megfelelően alapvetően kétféle adatszolgáltatói kör teszi ki a célsokaság egészét:

- **Gazdasági szervezet:** mezőgazdasági tevékenységet folytató jogi vagy nem jogi személyiségű gazdasági társaság, az egyéni vállalkozók és az egyéni gazdasági tevékenységet folytatók nélkül.

---

<sup>6</sup> Az Európai Parlament és a Tanács 1166/2008/EK rendelete (2008. november 19.) a gazdaságszerkezeti felmérésekről és a mezőgazdasági termelési módszereket vizsgáló felmérésről, valamint az 571/88/EGK tanácsi rendelet hatályon kívül helyezéséről.

- **Egyéni gazdaság:** a gazdaságküszöböt elérő mezőgazdasági tevékenységet folytató háztartás és az adószámmal rendelkező egyéni vállalkozás által működtetett gazdaság.

Az egyéni gazdaságok esetében a gazdaságküszöb fogalma bővebb kifejtést érdemel. A megfigyelt – meg nem figyelt gazdaságok (háztartások) közötti képzeletbeli határvonal meghúzásakor a statisztikusnak a mindenkori adatfelhasználói igényeket kell figyelembe vennie. Esetünkben a legerősebb felhasználói igényt a gazdaságszerkezeti összeírásokat szabályozó Európai Parlament és a Tanács 1166/2008/EK rendelete jelenti. Ez kimondja, hogy a tagországoknak a felmérési küszöbértéket úgy kell megválasztaniuk, hogy csak legfeljebb a teljes mezőgazdasági területük, illetve teljes számosállatban <sup>7</sup> mért állatállományuk 2 százaléka maradjon lefedetlen, és emellett azokat a gazdaságokat sem szabad kihagyni, amelyek terményeik bizonyos hányadát eladásra szánják. Ezt megelőzően az 571/88/EGK<sup>8</sup> rendelet még ennél is szigorúbb lefedettségi szabályokat támasztott, hiszen az eladásra termelésen túl az volt a követelmény, hogy a gazdaságküszöböt úgy kell meghatározni, hogy legfeljebb az ország teljes standard fedezeti hozzájárulásának<sup>9</sup> (SFH) 1 százaléka maradhat csak ki a megfigyelési körből. Ennek megfelelően a korábban kialakult, meglehetősen alacsony gazdaságküszöb, ami több évtizeden át meghatározta az összeírandó gazdaságok körét (lásd 1. táblázat), továbbra is megmaradt. A küszöbérték elemei között természetesen vagylagos kapcsolat van. Tehát ha egy háztartás rendelkezik legalább egy négylábú haszonállattal (sertés, szarvasmarha, ló vagy juh), akkor máris egyéni gazdaságnak minősül. A legutolsó, 2007-es Gazdaságszerkezeti Összeírás (GSZÖ) során is ezt alkalmazták.

---

<sup>7</sup> Számosállat: 500 kg élőszűlyű állatra átszámított egyenértékes.

<sup>8</sup> A Tanács 571/88/EGK rendelete (1988. február 29.) a mezőgazdasági üzemek szerkezetére vonatkozó, 1988. és 1997. évek közötti közösségi felmérések szervezéséről.

<sup>9</sup> Standard fedezeti hozzájárulás (SFH): az egyes jellemző mezőgazdasági tevékenységek esetében, adott régióban az átlagos helyzetnek megfelelő bruttó árás értéke, amely a bruttó termelési érték és a közvetlen változó költségek különbsége. A termék szintű kvócienseket az Agrárgazdasági Kutató Intézetben kalkulálják, a Mezőgazdasági Számviteli Információs Hálózat (FADN) a mezőgazdasági üzemek pénzügyi-, vagyoni helyzetét felmérő Európai Unió reprezentatív információs rendszer magyarországi alrendszere, a Tesztüzemi Információs Hálózat, ismertebb nevén tesztüzemi rendszer adatai alapján. (A mezőgazdaság GDP-szerű mutatója.)

1. TÁBLÁZAT: A GAZDASÁGKÜSZÖB ALAKULÁSA A TELJES KÖRŰ ÖSSZEÍRÁSOK SORÁN

Megnevezés	1972	1981	1991	2000	2010
Összes termőterület (hektár)	0,15	0,15	0,15	0,15	0,15
Gyümölcsös, szőlő, kert (hektár)	0,08	0,08	0,08	0,05	0,05
Védőtakarás alatti termőterület (hektár)					0,01
Szarvasmarha (darab)	1	1	1	1	1
Sertés (darab)	1	1	1	1	1
Ló (darab)	1	1	1	1	1
Juh (darab)	1	1	1	1	1
Kecske (darab)	-	-	-	-	1
Bivaly (darab)	-	-	-	-	1
Strucc (darab)	-	-	-	-	1
Baromfi (darab)	50	50	50	50	50
Méhcsalád (darab)	20	20	25	5	5
Nyúl (darab)	20	20	25	25	25
Egyéb kisállat (darab)	-	-	25	25	25 <sup>10</sup>

*Forrás: Laczka-Szabó (2000), 231. o. alapján saját szerkesztés*

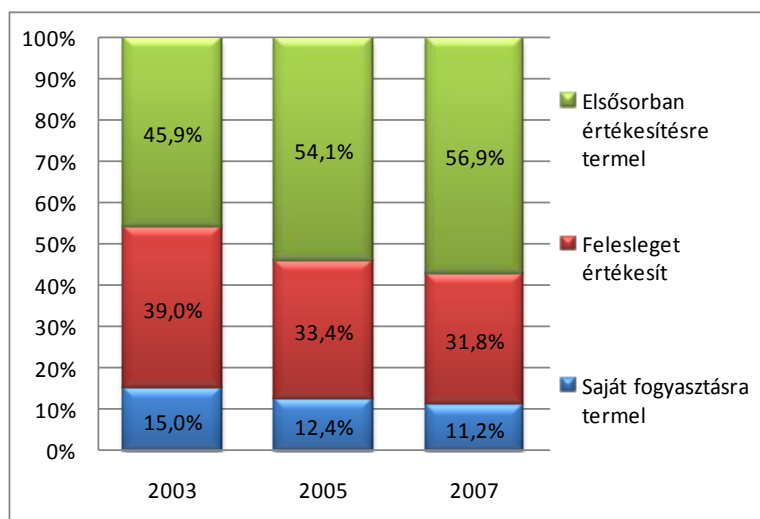
A 2010-es ÁMÖ esetében – ahogy az 1. táblázatból is kiolvasható – bekerült három új állatfaj a gazdaságküszöb definíciójába. A census során az a háztartás is egyéni gazdaságnak minősül, aki legalább 1 kecskét, bivalyt vagy struccot tart. Ez lényegében a nagyobb testű állatok körének bővülését jelenti. A kecske esetében erre azért lehetett szükség, mert bár Magyarországon nem kifejezetten jellemző a kecsketartás, de az EU-s előírások külön hangsúly fektetnek erre az állatfajra is. Az alacsony küszöbérték meghatározásában döntő szerepe van annak is, hogy Magyarországon jellemző volt a kisebb volumenű, háztáji, a saját és szélesebb értelemben vett család ellátására történő gazdálkodás. Ez csökkenő mértékben, de még napjainkban is fontos szerepet játszik bizonyos térségek, és az idősebb korosztály körében.

A gazdaságküszöböt azért szükséges ilyen módon meghatározni, hogy az összeíró a helyszínen viszonylag egyszerűen el tudja dönteni az adott háztartásról, hogy megfelel-e a küszöbértéknek. Valószínűleg árnyaltabb megoldás lenne valamilyen értékmutató alapján meghúzni a választvonalat, de ezek számítása a helyszínen, jelen körülmények között (papír alapú összeírás) túlságosan bonyolult lenne. Ráadásul az értéken alapuló mutatókhoz

<sup>10</sup> Az egyéb kisállatok alatt 2010-ben konkrétan a prémes állatokat és a húsgalambot értik.

használható koeficiensek sosem aktuálisan állnak rendelkezés, így az is ellenérv, hogy csak 2-3 éves szorzószámokat lehetne felhasználni az összeírások lebonyolítása során.

7. ÁBRA: AZ EGYÉNI GAZDASÁGOK SFH ÉRTÉKÉNEK MEGOSZTLÁSA A GAZDÁLKODÁS CÉLJA SZERINT, 2003-2005-2007



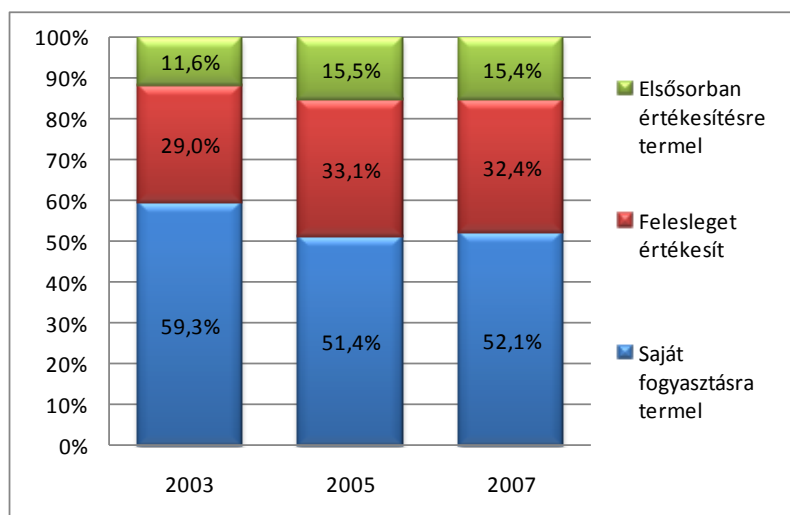
*Forrás: KSH adatbázis alapján saját számítás*

Gyakran éri kritika a mezőgazdasági összeírások során alkalmazott gazdaságküszöböt. Sokan úgy vélik, az túlságosan alacsony, és ilyen alacsony szinten nem is beszélhetünk mezőgazdasági termelésről. Valóban jelentős a csak saját fogyasztásra termelő egyéni gazdaságok száma (7. ábra) az alacsony küszöbértékek miatt, egész pontosan az egyéni gazdasági kör több mint felét ők teszik ki. Ez a nagy tömegű adatszolgáltató természetesen megnehezíti a gazdaságok összeírását és nyilvántartását minden szempontból. Ugyanakkor nem szabad megfeledkeznünk arról, hogy az SFH értékben mért teljesítménye ennek a körnek is viszonylag nagy jelentőséggel bír. Az előző időszakban lezajlott három gazdaságszerkezeti összeírásából jól látszik, hogy 10-15 százalékos arányt képviselnek az egyéni gazdaságok csoportján belül (lásd 8. ábra), ami nem elhanyagolható nagyságrend az EU-s jogszabályok előírásait is szem előtt tartva (a témában lásd még: Oros 1992; Galambosné Tiszberger 2009).

Ezek a tények támasztják alá leginkább, hogy miért is olyan fontos az egyéni gazdaságok megfigyelése, a megfelelő minőségű adatok összegyűjtése. A pontos és részletes mezőgazdasági statisztikai adatok fontosságát és szükségszerűségét ezeken túlmenően az is alátámasztja, hogy a racionálisan feltételezhető erőforrás felhasználási arányok (munkaerő,

föld, termelési eszközök) sok esetben nem érvényesülnek a gyakorlatban. Így ezek bemutatásához a gazdaságok szintjének megfigyeléséből kell kiindulni. (Oros 2002)

8. ÁBRA: AZ EGYÉNI GAZDASÁGOK SZÁMÁNAK MEGOSZTLÁSA A GAZDÁLKODÁS CÉLJA SZERINT, 2003-2005-2007



Forrás: KSH adatbázis alapján saját számítás

Ennek a szükséges, de viszonylag alacsony gazdaságkülönbnek, illetve a magyar mezőgazdasági tevékenység jellegzetességeinek köszönhetően az egyéni gazdaságok száma ugyan csökkenő tendenciát mutat, de így is rendkívül magas. A termelőszövetkezetek felbomlásával csaknem 2 millió háztartás foglalkozott számottevő mezőgazdasági termeléssel. Még a 2000. évi Általános Mezőgazdasági Összeírás során is közel 1 millió egyéni gazdaságot (958 534) találtak meg az összeírók a házról-házra történő felkeresések alkalmával. A tavalyi évben zajló census viszont már azt mutatja, hogy az elmúlt 10 évben az egyéni gazdaságok száma jelentős mértékben lecsökkent (40 százalékkal), már csak 567,5 ezer egyéni gazdaságot regisztráltak. (Lásd 2. táblázat.) A háztartások szintjén megjelenő mezőgazdasági tevékenység tehát egyértelműen visszaszorulóban van, ugyanakkor a több mint félmillió célsokaság statisztikai szempontból még mindig rendkívül nagy számosságúnak mondható.

Az egyes gazdaságtípusok teljesítményét is érdemes megvizsgálni, összehasonlítani. A 3. táblázat arányaiban szemlélteti a két gazdasági kör különbségeit. A gazdasági szervezetek elenyészőnek tűnő számossága mellé viszonylag magas termelékenység, hatékonyság társul. Az egy gazdaságra jutó SFH érték több tízezerszeresét állítja elő egy gazdasági szervezet. Ez



nyilvánvalóan tükröződik abban a tényben is, hogy a teljes fedezeti hozzájáruláson közel 50-50 százalékban osztozik a két gazdálkodói kör. A koncentráció mértéke tehát rendkívül jelentős az ágazatban, legalábbis, ami a statisztikai értelemben vett koncentrációt jelenti. Mezőgazdasági szempontból ugyanakkor nemzetközi szintre is kitekintve hatékonyságról és koncentrációról nem igazán beszélhetünk. (Buday-Sántha 2009) A holland vagy dán agrárágazat teljesítményéhez képest a magyar mezőgazdasági szervezetek jóval gyengébb képet mutatnak, mind a termelékenység, a koncentráció és a technológiai színvonal alapján.

2. TÁBLÁZAT: A GAZDASÁGOK SZÁMÁNAK ÉS SFH ÉRTÉKÉNEK ALAKULÁSA, 1972–2010<sup>11</sup>

Év	Egyéni gazdaságok			Gazdasági szervezetek		
	száma (ezer db)	SFH érték (millió Ft)	1 gazdaságra jutó SFH érték (ezer Ft/gazdaság)	száma (db)	SFH érték (millió Ft)	1 gazdaságra jutó SFH érték (ezer Ft/gazdaság)
1972	1 841,5	-	-	6 113	-	-
1981	1 529,6	-	-	1 449	-	-
1991	1 395,8	-	-	3 077	-	-
2000	958,5	316 922	330,63	8 382	248 160	565 082
2003	765,6	293 366	383,18	7 813	247 417	540 784
2005	706,9	274 318	388,06	7 897	247 114	521 432
2007	618,7	245 650	397,06	7 656	246 967	492 617
2010	567,5	..	..	8 800	..	..

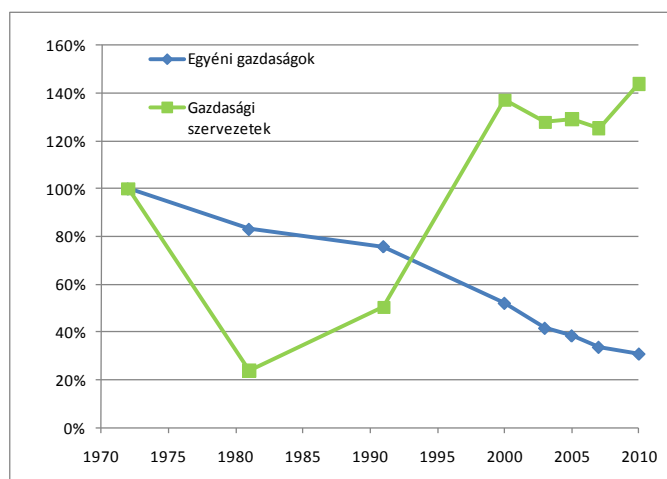
*Forrás: Általános Mezőgazdasági Összeírások, Gazdaságszerkezeti Összeírások alapján (www.ksh.hu). Saját szerkesztés.*

A két adatszolgáltatói kör elkülönítésének – azon túl, hogy jellemzőikben és a tevékenység volumene alapján lényegesen különböznek – az eltérő összeírás-technika és a rendelkezésre álló regiszterinformációk jellege miatt van alapja. Információ az egyéni gazdaságok mindegyikéről csak a censusok alkalmával kerül begyűjtésre. A 10 éves rendszerességgel zajló teljes körű összeírások között azonban az egyéni gazdaságok név- és címadatai frissítetlenek maradnak (kivéve azokat, akik valamilyen összeírás során a mintába kerülnek). A nehezebb elérhetőségük és a nagy számosságuk miatt az egyéni gazdaságokat csak minta alapján, összeírók segítségével figyelik meg.

<sup>11</sup> A 2000-nél korábbi időszakokban még nem használták az SFH mutatót, ezért ott ezek az adatok hiányoznak. 2010-től pedig az ún. standard termelési érték váltotta fel az SFH mutatót, ami tartalmát tekintve a kibocsátási értékre koncentrál, ezért lényegesen eltér a korábbtól. Ezért ezt nem tartottam célszerűnek feltüntetni. A két gazdálkodói kör az új mutató szerint is szinte pontosan fele-fele arányban járul hozzá a mezőgazdasági termelési értékhez.

A gazdasági szervezetek ezzel szemben a cégjegyzékben szerepelnek, a teljes kör azonosítása viszonylag egyszerű (azért a főtevékenység, melléktevékenység meghatározása regiszterinformációk alapján mégsem mindig egyértelmű). A gazdasági szervezetek neve és címe pontosan rendelkezésre áll, így a postai adatgyűjtés lehetősége megalapozott. Általános feltételezés pedig, hogy a szervezetek rendelkeznek az esetleg bonyolultabb kérdőívek kitöltéséhez szükséges „szakértelemmel” is. A be nem érkező kérdőívek begyűjtésére is több lehetőség van az egyéni gazdaságokhoz képest. Természetesen számosságuk is viszonylag alacsony, az elmúlt évtizedben 8 ezer körül mozgott (2. táblázat). A legtöbb mezőgazdasággal foglalkozó szervezetet a 2010-es census során regisztrálták. Számuk így már megközelíti a 9 ezret.

9. ÁBRA: A GAZDASÁGOK SZÁMÁNAK ALAKULÁSA, 1972-2010 (1972=100%)



*Forrás: Általános Mezőgazdasági Összeírások, Gazdaságszerkezeti Összeírások alapján (www.ksh.hu). Saját számítás és szerkesztés.*

A külön kezelt gazdaságcsoportok számosságának időbeli változása is nagyon eltérő képet mutat (lásd 9. ábra). Az egyéni gazdaságok száma folyamatos, nagymértékű fogyást mutat az elmúlt 40 évben, ami – legalábbis az utóbbi 2 évtizedben – a háztáji gazdálkodás egyértelmű visszaszorulásával magyarázható. A gazdasági szervezetek száma sokkal változatosabb, elsősorban a szocialista nagyüzemek felbomlása miatt. Az utóbbi 10 évben ugyanakkor viszonylagos stabilitást figyelhetünk meg.

A két adatszolgáltatói kör nemcsak számosságában, de az általuk létrehozott érték tekintetében is lényeges különbségeket mutat (3. táblázat). Az egy gazdaságra számított SFH értékek alapján is nyilvánvaló, hogy a gazdasági szervezetek sokkal koncentráltabb, nagyobb

volumenű, jobb hatásfokú termelést valósítanak meg, mind az állattenyésztés, mind a növénytermesztés területén. Gyakorlatilag több mint 1000-szeres szorzóval jutunk el egy egyéni gazdaság átlagos teljesítményéből egy gazdasági szervezetéhez. Ez az elaprózódás kelti azt az érzést a statisztikusban, hogy egyébként ebben a kicsi átlagos értékben ejtett statisztikai hiba kevésbé súlyosnak tűnik a nagy teljesítmény-értéket lefedő gazdasági szervezetek mellé helyezve. Ugyanakkor, a „sok kicsi sokra megy” alapon az egyéni gazdaságok mégis képesek Magyarország teljes mezőgazdasági termelésének közel a felét létrehozni. A mérleg mostanában látszik átfordulni a gazdasági szervezetek javára, de kerekítve még mindig azt mondhatjuk, hogy a két megfigyelési kör közel egyenlő arányban járul hozzá a végeredményhez.

3. TÁBLÁZAT: AZ EGYÉNI GAZDASÁGOK ÉS A GAZDASÁGI SZERVEZETEK ARÁNYELTOLÓDÁSAI, 2000–2007

Év	Egyéni gazdaságok			Gazdasági szervezetek		
	aránya (%)	SFH érték aránya (%)	átlagos SFH értéke a teljes gazdaságra jellemző átlag százalékában (%)	aránya (%)	SFH érték aránya (%)	átlagos SFH értéke a teljes gazdaságra jellemző átlag százalékában (%)
2000	99,1	56,1	56,6	0,9	43,9	96 688,2
2003	99,0	54,2	54,8	1,0	45,8	77 341,4
2005	98,9	52,6	53,2	1,1	47,4	71 479,7
2007	98,8	49,9	50,5	1,2	50,1	62 635,6

*Forrás: Általános Mezőgazdasági Összeírások, Gazdaságszerkezeti Összeírások alapján (www.ksh.hu). Saját szerkesztés.*

#### 2.4.2. Az összeírások rendszere

A Magyarországon hatályban lévő statisztikáról szóló törvény (1993. évi XLVI. törvény) általánosságban rendelkezik az adatgyűjtésekről. A rendszeres éves és évközi adatgyűjtéseket mind a gazdálkodó szervezetek, mind a természetes személyek gazdaságai esetében évente Kormányrendeletben elfogadott Országos Statisztikai Adatgyűjtési Program (OSAP) szabályozza. Az OSAP tartalmazza az adott évre a hivatalos statisztikai szolgálat szerveinek közvetlen adatgyűjtéseit és statisztikai célú adatátvételeit, valamint a gyűjthető adatköröket. Az OSAP minden évben kormányrendelet formájában jelenik meg. A kormányrendelet szabályozza a kérdőívek formai és tartalmi kellékeit. Minden kérdőív OSAP számmal rendelkezik. Ezt a nyilvántartási rendszert a KSH alakítja ki és vezeti.

A mezőgazdaságot leíró adatok tömegét több szervezet munkájának az eredménye adja. A mezőgazdasági adatok különböző szintű és típusú adatainak összegyűjtéséért és feldolgozásáért a Központi Statisztikai Hivatalon túl a következő szervezetek is felelősek: Földművelési és Vidékfejlesztési Minisztérium, Agrárgazdasági Kutató Intézet, Mezőgazdasági és Vidékfejlesztési Hivatal, stb.

A dolgozatban a KSH mezőgazdaság statisztikai rendszerének bemutatására fektetem a hangsúlyt, hiszen a mintavételi lehetőségek kutatása is ehhez kapcsolódik. Ezen belül is a termelésstatisztikai összeírásokra koncentrálok, hiszen a mintavétel szükségessége ezen a területen merül fel, az egyéni gazdaságok megfigyelésében. Emellett természetesen vannak egyéb témájú kérdőívek is; különböző szempontú árakra, ráfordításokra, szolgáltatásokra vonatkozóan (a teljes felsorolást lásd III. számú melléklet). A KSH mezőgazdasági adatgyűjtéseit azok gyakorisága, illetve jellege alapján 2 nagy csoportba sorolhatjuk. Az egyik, az évek során rendszeres, éves, illetve évközi összeírások rendszere. Ezt a 4. táblázat mutatja be szemléletesen. Egy év során tehát 8 jelentősebb témát feldolgozó, kifejezetten mezőgazdasági adatgyűjtés lebonyolítására kerül sor. A két leggyakoribb, a baromfikeltető állomások adatait összegyűjtő „Baromfikeltetés”, és a tej termelését, felvásárlását nyomon követő kérdőív, amelyeket havi rendszerességgel küld be az adatszolgáltatók (gazdasági szervezetek) teljes köre. 2009-től évente két alkalommal kerül sor az állatállomány számbavételére (június 1. és december 1. eszmei időpontokkal). Korábban három összeírás volt: április 1., augusztus 1. és december 1. eszmei időpontokkal, azonban az EU előírások lehetővé teszik a ritkább megfigyeléseket is, hiszen csak két időpontra vonatkozóan kell adatok szolgáltatnunk egy éven belül. A hazai igények ugyan megkövetelnék a három összeírást, de a költségek lefaragása erősebb szempontnak bizonyult. Ráadásul az áprilisi és augusztusi összeírások júniusival történő helyettesítésével lehetőség nyílt arra is, hogy a földterületi kérdőívet az állatállományával közösen kérdezzék le ugyanabban az időszakban. Az idősorban tulajdonképpen nem okozott problémát az áttérés, hiszen a decemberi referencia időpont minden időszakban állandó, így az összehasonlítás lehetősége megmarad. Az éven belüli szezonális ingadozások kimutatása, elemzése viszont lényegében ellehetetlenült.

Az állatállományt mind az egyéni gazdaságok, mind a gazdasági szervezetek szempontjából összeírják. Ugyanez a helyzet a májusi föld- és vetésterülettel, amit szintén mindkét adatszolgáltatói kör adatai alapján állítanak össze. Két, előzetes adatokat eredményező összeírás csak a gazdasági szervezetek körére fókuszál: a januári „Főbb

növénykultúrák terméseredménye”, és az augusztusi „Kalászos gabonák terméseredménye” című kérdőívek. Ezek feldolgozásával viszonylag gyorsan adathoz juthatnak a felhasználók a két témát tekintve. Természetesen az egyéni gazdaságok köre sem marad lefedetlen. Annak értékét szakértői becslés alapján állapítják meg, amelyhez különböző, egyéb forrásból rendelkezésre álló adatokat használnak fel (decemberi éves beszámoló, májusi föld- és vetésterület, stb.). A legnagyobb lélegzetű rendszeres összeírás az éves beszámoló. Ebben a naptári éves, teljes mezőgazdasági termelésről számot adnak az adatszolgáltatók. Mérlegszerű elszámolásokat is tartalmaz ez a kérdőív a növénytermelésről, állattenyésztésről, növényi- és állati termékekről, amelyek a mezőgazdasági számlarendszerhez biztosítanak alapvető információkat.

A rendszeres termelésstatisztikai összeírásokat bizonyos időszakonként nagyobb volumenű adatgyűjtések egészítik ki. A legnagyobb összeírás a mezőgazdasági census, amire a népszámlálásokhoz hasonlóan 10 évente kerül sor. A legutóbbi 2010-ben zajlott le. Ezek a censusok a FAO és az Eurostat előírásait egyaránt teljesítik.

4. TÁBLÁZAT: A RENDSZERES MEZŐGAZDASÁGI ÖSSZEÍRÁSOK

Időszak	Adatgyűjtés neve	Gazdasági szervezetek	Egyéni gazdaságok
január	Főbb növénykultúrák terméseredménye	×	
május	Föld- és vetésterület	×	×
	Állatállomány	×	×
augusztus	Kalászos gabonák terméseredménye	×	
december	Éves beszámoló	×	×
	Tej, tejtermékek	×	
havi	Tejtermelés, -felvásárlás	×	
	Baromfikeltetés	×	

*Forrás: Saját szerkesztés.*

Az Eurostat, az Európai Unió statisztikai hivatalának az a feladata, hogy uniós szinten harmonizált adatokat „állítson elő”, abból a célból, hogy segítsége a döntéshozatal folyamatát a közös politikák kialakításához, megvalósításához, valamint a tagországok teljesítményét összehasonlítsa. A mezőgazdaság területén sincs ez másképp, és mivel a Közös Agrárpolitika egy meglehetősen nagy „fejezet”, a támogatások rendszerében kifizetett összeg jelentős nagyságrendet képvisel (az EU költségvetésének közel 45 százaléka), különösen fontos az ágazatra vonatkozóan pontos, naprakész, harmonizált adatok megléte. Az Európai Unió információs rendszere több rétegből, több intézmény munkájából tevődik össze. A primer

információs rendszerhez tartoznak a nagy adatgyűjtők  
([www.ksh.hu/agrararcenzusok\\_amo\\_2010\\_nemzetkozi](http://www.ksh.hu/agrararcenzusok_amo_2010_nemzetkozi)):

- az agrárstatisztika, amely az Eurostat koordinálása mellett nyújt statisztikai információkat;
- a – korábban már említett – FADN, melynek feladata a gazdaságok szintjén jelentkező pénzügyi folyamatok, jövedelemhelyzet nyomon követése;
- a Piaci Információs Rendszer, amely kielégíti a brüsszeli apparátus információs igényeit, illetve tájékoztatja a termelőket a piaci folyamatokról;
- az Integrált Igazgatási és Ellenőrző Rendszer, amely a támogatások elnyerését, a kifizetések elszámolását és ellenőrzését végzi.

Emellett vannak szekunder információs rendszerek is, amely a primer adatbázisokból nyerik adataikat, és azokat dolgozzák fel. Ilyen többek között a mezőgazdasági számlák rendszere is. Ezek célja egy-egy „szűkebb” terület speciális információigényének biztosítása.

Az Eurostat-ról – bár az elsődleges információs rendszer része – fontos kiemelni azt is, hogy közvetlenül nem gyűjt adatokat, csak a tagországok által rendelkezésére bocsátott adatokat használja fel. Ezek képezik a tagországok összehasonlításának alapját, illetve az Európai Bizottság többi főigazgatóságának igény szerinti hozzáférését a harmonizált információhoz.

A mezőgazdaság statisztikában szigorú előírások vonatkoznak az évközi és a teljes körű összeírásokra egyaránt. Ezek azonban az EU előtt is léteztek, a hazai felhasználók igényei által vezérelve. Ami viszonylag új elem az a gazdaságszerkezeti összeírások rendszere. Ezek a censzusok közötti időszakban 2-3 alkalommal kerülnek lebonyolításra, kifejezetten EU-s előírás alapján, de nem csak EU-s felhasználásra, természetesen. Az évközi összeírásokhoz képest viszonylag nagy a mintavételi arány, és a megvizsgált mutatók köre is jóval szélesebb. Leginkább a decemberi éves beszámolóhoz hasonlítanak ezek a kérdőívek, de még annál is részletesebbek. Ennek megfelelően a legutóbbi 3 GSZÖ (2003, 2005 és 2007) a decemberi

rendszeres összeírással került összevonásra, annak az adatlapját egészítették ki a további szükséges kérdésekkel, mutatókkal. A szerkezeti összeírásokról egységes szerkezetben kell a mikroadatokat<sup>12</sup> tartalmazó adatállományt az Eurostat részére megküldeni. Ez az úgynevezett Eurofarm adatbázis vagy adatbank. Ez 250-350 mutatót (oszlopot) tartalmaz, amelyekre szigorú előírások vonatkoznak. Számos formai és tartalmi ellenőrzésen, belső konzisztencián alapuló szempontnak kell tökéletesen megfelelnie az átadott állománynak az elfogadáshoz. A mutatók definícióit, az ellenőrzési szabályokat, a formai követelményeket külön kézikönyv tartalmazza, szabályozza. Ez már az összeírás szervezésének időszakában rendelkezésre áll, így a kérdőív véglegesítéséhez az itt felmerülő igényeket is pontosan lehet követni. Az Eurofarm adatállomány elfogadása pedig alapját képezi az Európai Unió pénzügyi hozzájárulás megszerzésének. Ennek az állománynak az előállítása tehát nemcsak széleskörű, egységes és rendkívül gazdag adattartalma miatt fontos, hanem finanszírozási szempontból is. Ennek az adatszolgáltatási formának több előnye is van az Európai Unió adatigények szempontjából. Egyrészt nagyon jó minőségű adatokat kapnak meg a szigorú és mindenre kiterjedő ellenőrzési folyamatoknak köszönhetően. Másrészt az egységes szabályozásnak megvan az az előnye, hogy minden tagállam adatai ugyanabban a formában, ugyanazokat a mutatókat tartalmazzák, ami megalapozza az összehasonlítások és aggregálások elvégzését. Harmadrészt pedig az Európai Bizottság ad hoc adatigényei az alapadatok szintjéről nagyon gyorsan kalkulálhatóak, a nemzeti statisztikai hivatalok közbeiktatása nélkül. (Tájékoztatást utólag minden adatkérésről küldenek a tagországoknak is.) Ezeket a számításokat az teszi lehetővé, hogy külön oszlopban szerepelnek a „felszorzáshoz” szükséges koefficiensek is az adatállományban. Azok pontos felhasználásához pedig az Eurofarm mellé megküldött módszertani jelentések nyújtanak segítséget.

További tematikus összeírások is gazdagítják, kiegészítik a rendszeres adatgyűjtések eredményeit. Ilyenek a történeti áttekintésben már bemutatott szőlő- és gyümölcsös ültetvényekre vonatkozó összeírások. 2001-ben kerülhetett sor egy teljes körű szőlő- és gyümölcsös ültetvény összeírásra, ami a csatlakozási tárgyalásokhoz biztosította a szükséges információt. Emellett ez azért is jelentős volt, mert legutóbb 40 évvel korábban került sor az

---

<sup>12</sup> A mikroadat és az egyedi adat nem azonos fogalmak. Az egyedi adat azonosításra alkalmas formában áll rendelkezésre (névvel, címmel, stb.). A mikroadat fiktív azonosító sorszámmal ellátott adat, ami ugyan egy adatszolgáltatóra vonatkozik, de közvetlen azonosításra elméletileg alkalmatlan.

utolsó ilyen adatgyűjtésre. 2009-ben újra összeírták a szőlő ültetvényeket. 2007-ben 4 kiemelt gyümölcsre vonatkozóan került sor jelentősebb összeírás végrehajtására Magyarországon. Az alma-, körte-, őszibarack-, kajszibarack-ültetvények felmérést EU-s előírások tették kötelezővé a tagországok számára.

A mintavételes összeírások sorában még említést kell tenni a Mezőgazdasági háztartások jövedelem-felvételéről is. Ez a téma eddig érintetlen volt a magyar mezőgazdaság statisztikában. Eddig csak kísérletek, próbálkozások voltak a mintavétel és a lebonyolítás módszerének kidolgozására az évtized elején, de tényleges összeírás még nem zajlott le. Egyrészt jelentős az igény ilyen jellegű adatokra, hiszen az ebből származó információ a mezőgazdasági számlák összeállításához releváns alapot jelenthetne. Másrészt ugyanakkor ez a kényes téma (jövedelmek) igen érzékenyen érinti az adatszolgáltatókat. Az adatigény mellett tehát azt is mérlegelni kell, hogy megéri-e az ilyen nehezen mérhető, és esetleg nem pontosan bevallott adatok begyűjtését megszervezni, finanszírozni. A feldolgozott információ megbízhatósága és valóságosága több szempontból is megkérdőjelezhető lenne. Emellett rossz fényt vethetne a többi mezőgazdasági adatgyűjtésre is, bizalmatlanságot válthatna ki az adatszolgáltatókból.

A fenti bemutatás jól tükrözi, hogy mennyire sokrétű és komplex a mezőgazdaság statisztika rendszere. A nagyobb összeírások nélkül is rengeteg adat begyűjtésére, rögzítésére, feldolgozására és publikálására kerül sor éven belül is. Az adatgyűjtések részletei sok szempontból kötött pályán mozognak, ugyanakkor nagy szükség van a statisztikusok szakismeretére, lelkiismeretes munkájára a folyamatos adatigények jó minőségű kielégítésére.

Az Eurostat a jogszabályi előírásokon keresztül többek között szigorú minőségi elvárásokat is megfogalmaz. A dolgozat szempontjából elsősorban a becslési hibák nagyságára vonatkozó előírások fontosak, hiszen ezek teljesítéséhez, a pontosságnak való megfeleléshez keres a szerző hatékony módszertani megoldásokat.



### 3. Mintavétel elmélet

A dolgozat egyik alapvető célja, hogy hatékony mintavételi módszert találjon a magyarországi állatállomány összeírásához. Ennek a célnak az elméleti háttérét a mintavételi technikák, lehetőségek képezik. Ebben a fejezetben a főbb módszerek áttekintése mellett helyet kapott a mezőgazdaság statisztika gyakorlatában a közelmúltban és napjainkban felhasznált mintavételi technikák áttekintése is. Az adatok statisztikai minősége ugyan csak részben kapcsolódik a mintavételes összeírásokhoz, a teljesség igénye miatt azonban ezt is fontosnak tartom legalább nagyvonalakban áttekinteni, így a fejezet ezzel a témával zárul.

A statisztika módszertudományt csoportosítani tudjuk abból a szempontból, hogy milyen típusú adatállományból indulnak ki a számítások. Amennyiben a teljes alapsokaságra vonatkozóan nincsen információnk, csak annak egy részére, akkor következtetési statisztikáról beszélhetünk. (Részben a statisztikai döntéselmélet is ide tartozik.) Ennek két ága a becslésmélet és a hipotézisvizsgálat. (Pintér-Rappai 2007) Ezen módszerek mögött tehát egy valamilyen módon kiválasztott és megfigyelt mintasokaság adatai állnak. (Párniczky 1954)

A mintavétel elméleti megalapozása a matematikai-statisztika talán legfontosabb vívmánya a gazdasági-társadalmi jelenségek megfigyelése, számbavétele terén. (A mintavétel elmélet fejlődéséről lásd: Kish 1995a) A statisztika tárgya a tömegjelenségek vizsgálata. Ugyanakkor a tömeget nyilvánvalóan nem lehet vagy legalábbis általában nem kifizetődő, nem célszerű teljes egészében megfigyelni. A mezőgazdaságban ma Magyarországon tevékenykedő több mint 567 ezer egyéni gazdaságot sem célszerű minden adatgyűjtés alkalmával felkeresni és összeírni. A teljes sokaság megfigyelése gyakran nagyon költségigényes és hosszú időt igénylő folyamat. Éppen ezért rendkívül fontos, hogy a sokaság egy megfelelően kiválasztott, viszonylag kis részéből is megbecsülhetőek legyenek a kérdéses mutatók. Természetesen a kevesebb megfigyelésnek hátránya, ára is van. Ezt hívjuk mintavételi vagy becslési hibának. Azonban a bizonytalanság is számszerűsíthető matematikai módszerekkel, tetszőlegesen kiválasztott megbízhatósági szint mellett. Így a hibázás mértékének is tudatában lehetünk.

### ***3.1. Főbb mintavételi módszerek***

A dolgozat és a téma szempontjából legfontosabb, alapvető véletlen mintavételi módszereket veszem sorra részletesebben: egyszerű véletlen, rétegzett, csoportos, többlépcsős. Végül a nem véletlenül alapuló eljárásokat csak a teljességre való törekvés miatt említem meg.

#### ***3.1.1. Egyszerű véletlen mintavétel***

„A véletlen kiválasztás vagy véletlen mintavétel azt jelenti, hogy a minta kiválasztásakor biztosítjuk azt, hogy minden sokasági elem előre meghatározott – megtervezett – valószínűséggel kerüljön be a mintába.” (Hunyadi 2002 265. p.) A véletlen mintát valószínűségi mintának is szokták nevezni. Az egyszerű véletlen mintavétel definíciója Kish (1995b) alapján ennél annyiban szigorúbb, hogy az előre ismert kiválasztási arány minden mintavételi egység esetében egyenlő, vagyis bármely „n” elemű minta kiválasztásának valószínűsége megegyezik.

Az egyszerű véletlen mintavételnek előfeltétele, hogy rendelkezésünkre álljon egy olyan lista, amelyen minden alapsokasági elem pontosan egyszer szerepel. A kiválasztást úgy is meg lehet oldani, hogy minden sokasági elemhez különböző véletlen számot generálunk. Ezek alapján sorba rendezzük a sokaságot, így egy véletlen sorrendet kapunk. Ekkor már akár szisztematikus kiválasztással is ki tudjuk jelölni a mintába kerülő elemeket. (A szisztematikus kiválasztást általában a nem véletlen mintavételi módszerek között említik, de a kiválasztás jellegét alapvetően a kiinduló sokaság rendezettsége határozza meg.) Ez azt jelenti, hogy a véletlen sorszám alapján sorba rendezett sokaságból kiválasztjuk az első „n” elemet („n” a mintaelemszám), vagy véletlen kezdőszámtól minden „N/n”-dik elem kerül a mintába („N” az alapsokaság elemszáma). Ezzel a módszerrel a visszatevés nélküli kiválasztás egy formáját valósítjuk meg, hiszen minden elem legfeljebb egyszer kerülhet bele a mintába. A minta elemszáma logikusan nem haladhatja meg az alapsokasági elemszámot ( $n \leq N$ ). A dolgozat során az egyszerű véletlen (EV) mintavételt ennek a visszatevés nélküli változatnak a megnevezésére fogom használni.

Az egyszerű véletlen mintavételnek elméleti szempontból jelentőséggel bíró másik változata a visszatevéses mintavétel. Ebben az esetben minden kiválasztott elem a

kiválasztását követően visszakerül a sokaságba, és esélye van rá, hogy a következő kiválasztások alkalmával újra belekerüljön a mintába. Ez úgynevezett független azonos eloszlású (FAE) mintát eredményez, ami rendkívül jó elméleti tulajdonságokkal rendelkezik. Ilyenkor fennállhat az, hogy a kiválasztott minta elemszáma nagyobb lesz, mint az alapsokaság számossága („n” bármekkora szám lehet „N”-től függetlenül). Gyakorlati oldalról kevésbé jelentős, de elméleti oldalról jó példája ennek az érme feldobás esete. Ekkor az alapsokaság két elemből áll: fej és írás. A feldobást ismételve minden egyes alkalommal bármelyik elem lehet felül. A feldobások száma fogja meghatározni a minta elemszámát, vagyis az a kettőtől függetlenül alakul ki. A statisztikai célú összeírások esetében a visszatevési módszer tulajdonképpen információvesztést hordoz magában a visszatevés nélkülihez képest. Ugyanolyan elemszám mellett ugyanis előfordulhat az, hogy nem „n”, hanem csak „n-k” („k” azon elemek száma, amelyek egynél többször szerepelnek a kiválasztott elemek között) különböző elemről kapunk adatokat. Vagyis „k” esetben már olyan elemet figyelünk meg, amelyet korábban már megismertünk. Egyrészt ez az oka annak, hogy a gyakorlatban ezt a típust nem használják. Másrészt pedig az, hogy bizonyos esetekben nem is lehet ezt megvalósítani. Ha a minőségellenőrzésekre gondolunk, akkor belátható, hogy egy termék ellenőrzése gyakran annak „elhasználásával”, tehát megsemmisítésével jár együtt. Ilyenkor nyilvánvalóan nem lehet ugyanazt az elemet visszatenni az alapsokaságba és újra kiválasztani.

A gyakorlatban ritkán alkalmaznak tisztán egyszerű véletlen, még ritkábban a visszatevési egyszerű véletlen mintavételt. Ennek egyik oka lehet az, hogy földrajzilag nagyon szétszórt mintát eredményez. Másik oka lehet, hogy sokkal hatékonyabb mintavételi terveket is találunk ennél. Ennek ellenére nagyon fontos ez az alapvető módszer, több okból is. Egyrészt a viszonylag egyszerű matematikai háttére miatt. A legtöbb formulát a végtelen sokaságból vett visszatevési véletlen mintavétel esetében lehet a legkönnyebben felírni, követni és megérteni. Másrészt a legtöbb, gyakorlatban is alkalmazott mintavételi módszer felfogható úgy, mint az egyszerű véletlen szigorított változata, illetve visszavezethető az egyszerű véletlenre. Végül pedig a gyakorlati munka során is sokszor hasonlítjuk össze a bonyolultabb mintavételi technikákkal elérhető hatékonyságot az EV mintáéhoz, ami egy jó kiindulási alapul szolgál. A dolgozatban magam is végzek ilyen összehasonlításokat a későbbiekben.

### 3.1.2. Rétegzett mintavétel

A rétegzés a sokaság egy vagy több alkalmasan megválasztott szempont szerinti csoportosítását jelenti. Az így kialakult rétegeken belül pedig általában valamilyen véletlen kiválasztást hajtunk végre. (Ebben különbözik a csoportos kiválasztástól. Csoportos mintavétel esetén ugyanis a csoportokból – elsődleges mintavételi egységből – választunk véletlen mintát, és a kiválasztott csoportokon belül figyelünk meg minden elemet – végső mintavételi egységet.) Ennél a gondolatnál rögtön meg is állnék. Ez az általános megfogalmazás már magában rejti a rétegzett mintavétel további csoportosítási lehetőségét. Eszerint ugyanis maga az ismerv, illetve a rétegzés történhet „bárhogyan”. Általánosságban tehát nem mondunk semmit a rétegző ismérvek tulajdonságairól. Nyilván szükség van arra, hogy a felhasználásra szánt rétegző ismerv(ek) szempontjából előzetes információval rendelkezünk, és az alapján minden sokasági elemet el tudjunk helyezni, be tudjuk sorolni. A rétegzést a csoportosításhoz hasonlóan, egyértelműen kell elvégezni, vagyis minden elemet be kell sorolni pontosan egy rétegbe, illetve egy elem nem tartozhat egyszerre több réteghez. A teljes sokaság nagysága ebben a szemléletben így írható fel<sup>13</sup>:

$$N = N_1 + N_2 + \dots + N_L = \sum_{h=1}^L N_h$$

Felmerül a kérdés, hogy milyen tulajdonságokkal rendelkezik egy megfelelő rétegző ismerv. Ez az a pont, ahol a szakirodalom nem kezd el kategóriákat képezni, hanem kétféleképpen jár el. Az egyik irány az, amikor minden fajta rétegzést „egy kalap alá” vesznek (Ay 1976, Cochran 1977, Kish 1995b), nem törődve a rétegző ismérvek tulajdonságaival. A rétegzést, mint technikát, mint eljárást kezelik, és felsorolják, hogy egyébként a rétegzés milyen ismérvek mentén, milyen kiinduló feltételek mellett/miatt valósul meg. A másik elgondolás szerint csakis az számít rétegzésnek, amikor a becsülni kívánt változóval fennálló sztochasztikus kapcsolat alapján történik a rétegző ismerv kiválasztása, vagyis a sokaságot homogénebb rétegekre bontjuk a mintavétel előtt (Marton 1991, Hunyadi-Vita 2002, Pintér-Rappai 2007). Minden más esetet figyelmen kívül hagynak. Marton a területi rétegzést említi

---

<sup>13</sup> N – alapsokasági elemszám; h – rétegek; L – az összes réteg száma

ezen felül, de itt nem is rétegeknek nevezi a területi egységeket, hanem tartományoknak. Legtöbbször azonban szó sem esik más lehetőségekről.

A rétegeképítés alapjától függetlenül, módszertanilag ugyanúgy járunk el, mind a mintavétel, mind a becslések elkészítése során minden esetben. Biztosítjuk, hogy minden rétegből kerülnek be elemek a végső mintába. Ezért a külső szórás<sup>14</sup> gyakorlatilag elimináljuk, és csak a belső szórások<sup>15</sup> mértéke fogja befolyásolni a rétegzett mintából történő becslések hibáját (természetesen az adott mintavételi arány és az adott megbízhatósági szint mellett). A rétegeken belüli kiválasztás módszere, aránya lehet csak eltérő, de mindenképpen szerepel az összes réteg. A rétegzett mintavétel módszerének típusait, okait a következők szerint lehet összefoglalni (Galambosné Tiszberger 2011a):

1. A becsülni kívánt **változóval erős sztochasztikus kapcsolatban álló ismérvet** keresünk, és ez fogja jelenteni a rétegeképítés alapját. Egy ilyen ismerv megválasztása elsősorban a standard hibára gyakorolt jótékony hatása miatt előnyös. Ekkor a heterogén alapsokaságot viszonylag homogén (homogénebb) rétegekre bontjuk. A rétegzett mintavételből nyert adatokból történő becslés esetén a rétegek közötti eltérések hatását, mint már említettem, kiküszöböljük, és csak a belső szórás fog szerepet játszani a standard hiba számszerűsítése során. Tehát a külső szórás nagyságát már „megnyertük” a becslési hiba számszerűsítése során. Ráadásul a vizsgált mutató(k) belső szórását ebben az esetben viszonylag alacsonynak gondoljuk, hiszen éppen az volt a lényeg, hogy a rétegeképítő ismérvek szempontjából egymáshoz hasonló egyedek kerüljenek azonos rétegbe, így a teljes sokasághoz képest lényegesen homogénebb csoportokat kapunk. Ennek megfelelően az eliminálható külső szórás képvisel viszonylag magasabb arányt. Ebben az esetben tehát az **ismerv megválasztásával** hajtunk végre optimalizálást a mintavételi hiba nagyságának

---

<sup>14</sup> Külső szórás: A teljes csoportosított sokaságban a részátlagok főátlagtól való eltérésnégyzeteinek súlyozott átlaga  $\sigma_K = \sqrt{\frac{1}{n} \sum_{h=1}^L (\bar{x}_h - \bar{x})^2}$  (Pintér-Rappai 2007 237. o.)

<sup>15</sup> A belső szórás a teljes sokaságon belül az egyes értékek átlagos eltérését mutatja meg a saját csoportjuk részátlagától.  $\sigma_B = \sqrt{\frac{1}{n} \sum_{h=1}^L n_h \sigma_h^2}$  (Pintér-Rappai 2007 236. o.)

csökkentése érdekében. Végtelen esetben maga a becsülni kívánt jellemző lehetne a kiválasztott rétegeképző ismérv, de mivel erre irányul a felmérés, erről pontos információnk nyilvánvalóan nincsen, esetleg valamilyen közelítő ismerettel rendelkezhetünk erre vonatkozóan korábbi adatgyűjtésekből vagy esetleg adminisztratív adatforrásokból. Az esetek többségében inkább valamilyen segédváltozót, vagy változókat használunk fel a gyakorlatban.

2. Erős indoka lehet a rétegzésnek az **összeírási költségek optimalizálása**. Előfordulhat, hogy nagyon eltérő költségekkel lehet csak bizonyos részsokaságokat elérni, ezért eszerint is elvégezhető a rétegzés, és így a költségek bizonyos határok között maradhatnak azáltal, hogy a „drágább” elemekből kisebb mintával is megelégszünk. Tisztán költség alapú optimalizálás esetében tehát fel sem merül a homogenizálás igénye a mintavétel során.
3. A rétegzés módszerét azért is választhatjuk, mert a különböző rétegekbe tartozó egyedekre **másféle összeírási, megfigyelési módszert** szeretnénk vagy vagyunk kénytelenek alkalmazni. Ennek egyik oka az lehet, hogy fizikailag nagyon különböző helyen találjuk meg az egyes rétegekhez tartozó megfigyelési egységeket. Például egy multinacionális vállalat esetében valószínűleg más módszereket alkalmaznak egy dolgozói felmérés esetén az anyavállalatnál dolgozókra, illetve a leányvállalatok munkavállalóira, akik sok esetben nagyon távol vannak a központtól, és ebből adódóan akár eltérő munka-kulturális jegyekkel bírnak. Másik ilyen szempont lehet az, amikor az egyes sokasági elemek nagyon eltérő természetűek. Ezen azt értem, hogy másféle módszerrel lehet őket hatékonyan felmérni. Egy szellemi munkát végző ember valószínűleg könnyebben tölt ki egy kérdőívet, mint egy betanított fizikai munkás. Az utóbbi csoporttal lehetséges, hogy célravezetőbb interjút készíteni, vagy összeírói segítséget adni a kérdőív kérdéseinek megválaszolásához. Ilyen példát a mezőgazdaságból is hozhatunk, hiszen a gazdasági szervezetek, illetve a legnagyobb egyéni gazdaságok esetében feltételezzük, hogy rendelkeznek olyan szintű nyilvántartásokkal, könyveléssel, illetve szakértelemmel, ami alapján önállóan is ki tudják tölteni a postai úton megkapott kérdőíveket. (Természetesen a kérdőívvel együtt kiküldött kitöltési útmutató ebben segítséget nyújt.) Ugyanakkor az egyéni gazdaságok nagyobb részéhez, akik kisebb volumenben, kevesebb nyilvántartással dolgoznak, inkább összeírókat küldünk ki, akik segítenek a kérdőívet pontosan

kitölteni. Ilyen esetekben sem valószínű, hogy a rétegek a teljes alapsokaságon belül homogénebb részsokaságokat jelentenek, hiszen alapvetően nem vesszük figyelembe a vizsgált változókat a rétegek meghatározása során.

4. Okként merülhet fel az a helyzet, hogy **másféle információval rendelkezünk a különböző egyedekről**. A mezőgazdaság statisztika egyik nagy problémája az, hogy az egyéni gazdaságok azonosító adatairól nincsen egy naprakész regiszter. A censusok alkalmával kialakul egy teljes sokaságot lefedő nyilvántartás, de annak frissítése csak részben lehetséges két teljes körű összeírás között. A gazdasági szervezetekről ugyanakkor a cégnyilvántartás alapján viszonylag pontos, naprakész és teljes képet kapunk. A szervezeteknek tehát lehet postai úton küldeni a kérdőívet, hiszen pontos név és címadatok állnak rendelkezésre. Az egyéni gazdaságok esetében ez jelenleg nem megoldható, vagy legalábbis a visszaérkezési arányt lényegesen rontaná a kézbesítetlen levelek mennyisége. Ekkor is két rétegről beszélünk módszertani szempontból, viszont az összeírt változók tulajdonságait egyáltalán nem vizsgáljuk és nem használjuk ki.
5. Az is vezérelheti bizonyos rétegek külön kezelését, hogy az így képzett **részsokaságra vonatkozóan** is szeretnénk látni a **becslés** eredményeit, illetve ha külön előírások vannak bizonyos részsokaságokból elért pontosságra, hibanagyságra vonatkozóan. Másképpen fogalmazva, szeretnénk, ha bizonyos szempontok szerint reprezentatív lenne a mintánk. (Pintér-Rappai 2001) A magyar viszonyok között, és általában az EU többi országában is, ilyen rétegeket jelent a NUTS 2-es szintű területi beosztás. Magyarországon ez a régiók szintjét jelenti. Tehát a régiót már a kiinduláskor réteggépző ismérvnek tekintjük a legtöbb téma kapcsán, mert régiós adatokat is közölnek a Központi Statisztikai Hivatal kiadványai. Másképpen fogalmazva azt szeretnénk, hogy a minta a régiókra is legyen reprezentatív. Ezt a hazai felhasználói igények, illetve az EU-s szabályozás is szükségessé teszi. Esetenként még megyei szintű adatok előállítására is sor kerül a nemzeti szempontok alapján. Az angol nyelvű szakirodalomban egyébként külön elnevezést is kapnak az ilyen jellegű rétegek: domain. A magyar terminológiában ezt nem különítjük el. A legtöbb téma szempontjából a területi besorolás, rétegzés ugyancsak nem a homogenitás növelésének irányába mutat. Inkább az a jellemző, hogy az országos eloszlás

leképeződése látszik a területi szinteken is. Vannak persze kivételes témák, amelyek esetében a földrajzi hovatartozás homogenizál, de ezek előfordulása csak esetleges.

6. Egy speciális esetként fogható fel, amikor **ritka populáció**<sup>16</sup> megfigyelése a cél. Ilyenkor úgynevezett **aszimmetrikus** mintarétegzést hajthatunk végre. Ezt akkor célszerű alkalmazni, ha a ritka populáció rétegtképző ismérvek szerinti eloszlása ugyan ismeretlen, de azt tudjuk róla, hogy nem egyenletes. Azokat a rétegeket szükséges felülreprezentálni a mintavétel során, amelyekben a ritka populáció tagjainak előfordulása nagyobb valószínűséggel bír. Így jó eséllyel növelhető a hasznos minta elemszáma. (Kapitány 2010) Példa lehet erre a magas vérnyomásban szenvedők célsokasága, ahol a célszemélyek előfordulása a középkorú, illetve idősebb népesség körében valószínűleg nagyobb gyakorisággal bír. A vizsgált mutatók kérdése alapvetően itt sem merül fel. A rétegzés tényét és mikéntjét a célsokaság minél jobb elérése motiválja.

Az eddigi felsorolásból is látszik, hogy a rétegzett mintavétel okai és megvalósulási módozatai rendkívüli sokszínűséget mutatnak. Ráadásul az előbbi esetek szinte bármilyen kombinációja is előfordulhat a gyakorlatban, ami tovább bővíti a lehetőségek számosságát.

A rétegzés egyik hasznos eredménye az, hogy a rétegtképző ismérv(ek) szempontjából a sokaság minden fontos része be fog kerülni a mintába, így reprezentálni fogja az alapsokaságot ilyen értelemben. Ez általánosságban igaz. Emellett a fenti felsorolásból jól látható, hogy többféle indok, indokrendszer alapján juthatunk el a rétegzett mintavétel valamilyen formájának használatához. Azt tartom a legfontosabbnak, hogy a becslés hatásossága, pontossága szempontjából a gyakorlatban csakis az elsőként említett feltétel az, ami biztosan javítja a becsléshez tartozó hibaszámok alakulását az egyszerű véletlen kiválasztáshoz képest. Hiszen ebben az esetben **azért és csak azért** választjuk ki a rétegtképző ismérvet, mert az – célszerűen erős – sztochasztikus kapcsolatban áll a vizsgált változóval, és ezáltal homogénebb csoportokat hoz létre a megfigyelés célja szerint. Az összes többi (2-6. esetek) esetben más motiválja a rétegzést, és a homogénebb csoportok létrejötte csak

---

<sup>16</sup> Ritka populáció: olyan közösségek, társadalmi csoportok, amelyekről közvetlen mintavételi keret nem áll rendelkezésre. (Kapitány 2010)



esetleges, de semmiképpen sem törvényszerű velejáró. Ebből pedig az következik, hogy lényegét tekintve véleményem szerint külön alcsoportként kezelendő az első eset az összes többitől. Az első esetet **kapcsolati rétegzésnek** neveztem el, hiszen az ismerv és a vizsgált változó kapcsolata alapján dől el a rétegzés kérdése. A többi esetet összefoglalóan **technikai rétegzésnek** nevezem, mivel a mintavétel és a becslések elvégzése során a lépéseket és a módszertant tekintve úgy járunk el, mint az első esetben, csak éppen valószínűleg nem érünk el számottevő csökkenést a mintavételi hiba nagyságában (az egyszerű véletlen kiválasztáshoz képest). Azért is fontosnak tartom egy ilyen fajta elkülönítés bevezetését, mert a szakirodalom alapvetően azt sugallja, hogy a rétegzés a becslés hibáját javító művelet, a gyakorlatban viszont ez gyakran nem, vagy csak részben igaz. Látnunk kell tehát, hogy más okok, feltételek is eredményezhetnek rétegzett mintát, a hatékonyság csekély mértékű javítása mellett. Ezek alapján két részre bontva megfogalmazható a dolgozat első tézise:

### 1. tézis

**T1/A: A rétegzett mintavételt, mint gyakran alkalmazott mintavételi eszközt, a rétegző ismerv tulajdonságai, illetve a rétegzés okai, céljai szerint két alcsoportra bonthatjuk: kapcsolati és technikai rétegzésre.**

**T1/B: A kapcsolati rétegzés célja a becslések hatékonyságának javítása, míg technikai rétegzés esetén a rétegzés módszerét, csak mint osztályozó eszközt használjuk fel, de a becslések hatékonysága nem feltétlenül javul.**

A rétegzett mintavétel során a rétegekbe sorolt sokasági elemekből rétegenként egymástól függetlenül végezzük el a minta kiválasztását. Ez azt is jelenti, hogy az egyes rétegekben történő mintavétel egymástól független esemény. A kérdés az, hogy milyen megoszlással, és mekkora mintát válasszunk ki az egyes rétegekből. Természetesen a maximális mintaelemszámot meghatározhatjuk egyrészt a pénzügyi korlátokból vagy a pontossági követelmények és a sokasági szóródás (rétegenkénti belső szórások) segítségével. A rétegek között a mintaelemszámot alapvetően háromféle módon tudjuk elosztani. A háromféle módszer tárgyalásában az egyszerűbbtől haladok a bonyolultabb felé.

- **Egyenletes rétegzés:** Ebben az esetben a teljes sokaságra már valamilyen módszerrel meghatározott mintaelemszámot egyszerűen elosztjuk a rétegek számával, és minden rétegből ugyanakkora mintát veszünk ( $n_h = n/L$ , ahol  $L$  a rétegek száma). Ez a legegyszerűbb allokációs módszer. Nem igényel előzetes tervezést, és könnyű kivitelezni. Általános feltételek mellett, ha az egyes rétegek mutatóira is kíváncsiak vagyunk, szintén jó megoldás lehet ez a fajta allokáció. Akkor célszerű alkalmazni, ha hasonló méretű rétegeink vannak, ilyenkor az arányoshoz hasonló eredményeket ad.
- **Arányos rétegzés:** Amennyiben a rétegenkénti elemszámok nagyobb eltérést mutatnak, akkor az egyenletes elosztás helyett hatékonyabb a nagysággal arányos vagy egyszerűen csak arányos rétegzés használata. Ebben az esetben a rendelkezésre álló mintanagyságot az egyes rétegek nagyságával arányosan osztjuk szét a rétegek között. Ekkor a mintában a rétegek közötti megoszlás éppen az alapsokasági összetétellel fog megegyezni. Ez formálisan a következőt jelenti:  $n_h/n = N_h/N$ , vagyis  $n_h/N_h = n/N$ . Minden rétegben biztosítjuk ezáltal az azonos mintavételi arányt. A nagyobb rétegből nagyobb, a kisebb rétegekből pedig kisebb elemszámú minta kiválasztására kerül sor ezzel a módszerrel.
- **Optimális rétegzés:** A mintaelemszám elosztását más szempontok szerint is meghatározhatjuk. A két leggyakrabban alkalmazott optimalizálási szempont az egyes elemek összeírási költségei, illetve a rétegeken belüli szóródás. Bármelyiket is választjuk (kombinált alkalmazásuk is lehetséges), nyilván minimalizálás lesz a cél. Formalizáltan ez a következőt jelenti a rétegenkénti mintaelemszám meghatározásában:  $n_h = n \frac{N_h O_h}{\sum N_h O_h}$ . (Ahol  $O_h$  az optimalizálási tényező [ami lehet költségtényező, szórás] értéke a  $h$ -dik rétegben.) Ezt az allokációt akkor célszerű felhasználni, ha a rétegek között egyénként jelentős eltérések vannak az összeírási költségek vagy a szóródások (varianciák) tekintetében. A költségek oldaláról talán egyszerűbb megragadni a problémát, mert a költség a teljes kérdőívre, vagyis az összes változóra együttesen vonatkozik. Viszont a szóródás esetében ki kell választanunk egy „legfontosabb” vagy egy fontos és viszonylag nagy szórással jellemezhető mutatót, ami mentén az optimális allokációt elvégezzük. A szórásminimalizálás vezérelte optimális rétegzést annak 1934-ben első felhasználójáról Neyman-féle optimális rétegzésnek vagy Neyman-allokációnak is

szokták nevezni. (Később bukkantak csak rá, hogy Csuprov már 11 évvel korábban bemutatta a módszert (Cochran 1977).)

A rétegzés természetesen valamennyi mintavételi eljárással jól kombinálható, a gyakorlatban pedig nagyon elterjedt módszer.

### 3.1.3. *Csoportos mintavétel*

Viszonylag nagy sokasági elemszám esetén (több ezer, több százezer) akár a rétegzett, akár az egyszerű véletlen kiválasztás egymástól fizikailag nagyon különböző pontokon elhelyezkedő mintaelemeket adhat eredményül (a mintaelemek szétszórtnan helyezkednek el). Ilyenkor felmerül az a probléma, hogy egy ilyen szétszórtna mintát a távolságok miatt nagyon költséges, és esetleg időigényes is összeírni, megfigyelni. Amikor az egyes elemek megfigyelési költsége magas, akkor jöhet szóba a csoportos mintavétel.

A csoportos mintavétel előfeltétele az, hogy a sokaságot csoportokba tudjuk rendezni. Nagyon fontos, hogy minden elem csak egy csoporthoz tartozzon (ne legyen átfedés a csoportok között), ugyanakkor minden elemet besoroljuk egy csoportba (ne maradjon ki senki/semmi a csoportosításból), illetve, hogy egy-egy csoport több elemet tartalmazzon. A csoportokhoz tartozó elemek száma fogja megadni az egyes csoportok méretét. Ez a csoportosítás leggyakrabban fizikailag, területileg egymáshoz közel eső elemeket helyez azonos csoportba.

Matematikai-statisztikai megfontolásokból általánosságban elmondható, hogy előnyösebb, ha a csoportokon belül azonos mennyiségű végső mintavételi egység található, azonos méretűek a csoportok. Ez azonban a természetben és a társadalomban általában nem fordul elő. Példa lehet az egyforma méretű csoportokra a termékellenőrzések során az egy dobozban, kartonban, raklapon lévő elemek csoportja, ahol mesterségesen kialakított, egyforma mennyiségek találhatóak. A mezőgazdaság-statisztikában a csoportok kialakítását a címek körzetítésével oldották meg, amelyet általában a teljes körű összeírások előkészítéseként végeznek el a KSH területi szerveinél dolgozók. Az összeírási körzetek kialakításánál az a legfontosabb szempont, hogy a census lebonyolítása során egy összeíró számára a rendelkezésre álló idő alatt (2-3 hét) összeírható mennyiségű gazdaság kerüljön egy körzetbe. Az is kialakult „szabály”, hogy egy település minimum egy összeírási körzetet kell, hogy

jelentsen (vagyis hiába kicsi, kevés gazdaságok tartalmaz egy település, nem lehet másikkal összevonni). Az első feltétel az azonos, vagy legalábbis hasonló méretű csoportok kialakulásának irányába hat (bár az alföldi tanyavilágban a nagy távolságok miatt kevés gazdaság tartozik egy körzetbe), de a második feltétel gyakran hoz létre extrém alacsony elemszámú körzeteket is.

A csoportos vagy egylépcsős kiválasztás onnan kapta a nevét, hogy az elsődlegesen kiválasztott mintavételi elemeken (csoportokon) belül már minden végső mintavételi egység a mintába kerül. Vagyis nem egyenként választjuk ki a megfigyelés elemeit, hanem csoportosan, egész csoportokat kijelölve.

Csoportos kiválasztás esetén az alapsokaság tulajdonképpen elsődlegesen „M” darab csoportból áll, amelyből „m” számú mintaelemet választunk ki valamilyen véletlen kiválasztási elvet alkalmazva. A teljes mintanagyság tehát a csoportok méretétől is függ:

$$n = \sum_{j=1}^m N_j, \text{ azonos csoportméret esetén } n = m \times N_j$$

ahol m a csoportok számát jelöli.

A csoportok kiválasztását alapvetően háromféle módszerrel tehetjük meg:

- azonos valószínűséggel
- a csoport nagyságával ( $N_i$  vagy  $X_i$ ) arányos, eltérő valószínűséggel
- rétegzéssel.

(Az első két módszer ugyanazt az eredményt adja abban a speciális esetben, amikor a csoportok azonos nagyságúak.)

A megfigyelt változó nagyságával arányos kiválasztás elméletileg hatékony megoldásnak tűnik, és jól ki is használja az összeírás tárgyát képező ismerv tulajdonságait, de gyakorlati alkalmazhatósága olykor mégis akadályokba ütközik. Az összeírásokra – különösen a mezőgazdasági adatgyűjtésekre – a többcélúság jellemző. Vagyis többféle változó áll a vizsgálat középpontjában, ezért nemigen lehetséges egy „legfontosabb” változót kiválasztani, és annak a nagyságával arányosan elvégezni a csoportok kiválasztását. Illetve az is

meghiúsíthatja a módszer alkalmazását, hogy nem áll rendelkezésünkre információ a kiválasztott változó csoportok szerinti alakulásáról.

Csoportos mintavétel esetén a becslés hibája a teljes szórásból csak a külső szórást tartalmazza, vagyis „megnyerjük” a belső szórás mértékét. Ennek oka az, hogy mivel a csoporton belül minden végső megfigyelési egység bekerül a mintába, ezért a csoporton belül nem „követünk el” hibát. A szórás további csökkentésére is lehetőség van, úgy, hogy egyrészt nagyobb mintát veszünk a csoportokból, vagyis több csoportot választunk ki, vagy rétegezzük a csoportokat a kiválasztás előtt megfelelő szempontok alapján.

A csoportos mintavétel az egyszerű véletlen mintavételhez képest (ugyanakkora elemszám mellett) előnyösebb, illetve előnytelenebb tulajdonságokkal is bír. Hatékonysága általában rosszabb az egyszerű véletlen mintavételhez képest, hiszen a becslések standard hibája magasabb lesz a csoportokon belüli viszonylagos homogenitás következtében. (Ay 1976) Hunyadi (2001b) pontosan levezeti, hogy a csoportos mintavétel hatékonysága hogyan alakul az egyszerű véletlen kiválasztáshoz képest. Itt konkrétan kiderül, hogy minél nagyobb a csoportokon belüli homogenitás ( $RHO$  – Rate of Homogeneity), „annál nagyobb lesz a csoportos mintából történő értékösszegbecslés varianciája, mint a megfelelő EV mintából” (Hunyadi 2001b 44-45. oldal). Elméletileg azonban létezhet olyan eset is, amikor a csoportok heterogénebbek, mint általában a sokaság. Ekkor hatékonyabb lehet a csoportos felvétel a csoportosítás nélkülénél. Azt gondolom, és a tapasztalatok is azt sugallják, hogy a magyar mezőgazdaság területi alakulása nem felel meg ez utóbbi, kivételes eshetőségnek (ezért itt a csoportos mintavétel azonos feltételek mellett kevésbé hatékony, mint az EV kiválasztás). A csoportokon belüli homogenitás következtében a csoportos kiválasztás sok hasonló (esetleg felesleges) megfigyelést „hozhat”, ami a hatékonyság rovására mehet. A másik „kellemetlen” mellékhatása ennek a módszernek az, hogy a statisztikai becslések bonyolultabbá válnak. Ugyanakkor kétségtelenül alacsonyabbak lesznek az egy elemre jutó összeírási költségek, viszont az egyes elemek csoportba sorolása, a csoportosítás kialakítása esetenként többlet kiadást jelent. Az összeírást lényegesen gyorsabban lehet lebonyolítani a megfigyelési elemek közelsége miatt. Előnyös lehet az is, hogy bizonyos esetekben csak a kiválasztott csoportokban kell pontos lajstromot készíteni a sokaság egyedeiről, vagy még ez is elmaradhat (amennyiben a csoportosításhoz ez egyébként nem szükséges). Amikor a csoportos mintavétel kialakításának lehetőségét mérlegeljük, akkor ezeket a szempontokat kellene együttesen figyelembe venni, mérlegelni.

### 3.1.4. *Többlépcsős mintavétel*

A csoportos az egyszerű véletlen mintavételhez képest a becslőfüggvény magasabb varianciáját hozza magával, viszont alacsonyabb költségeket jelent. A csoportokon belül a további alcsoportok képzésének felhasználása a két egymással ellentétes tulajdonság közötti kompromisszum keresésének eredménye. A többlépcsős mintavétel tehát a csoportos kiválasztás továbbfejlesztett változata. Az első lépcsőben kiválasztott elsődleges mintavételi elemeken (primary sampling units - PSU) belül ugyanis nem kerül be minden egység automatikusan a mintába, hanem további kiválasztási eljárás kezdődik további alcsoportokon keresztül. Tehát második lépcsőben a már kiválasztott csoportokon belül alkalmazunk ismét valamilyen mintavételi módszert, technikát. Lehet ez véletlen mintavétel, illetve alkalmazhatunk rétegzést is. (Ebből a szemléletből látszik leginkább, hogy a csoportos mintavételt miért tekinthetjük egylépcsősnek.) Természetesen kettőnél több lépcső is elképzelhető, ha lehetőség van többféle szempontból egymásba ágyazott csoportokba rendezni a végső mintavételi elemeket.

Összességében több oka is lehet annak, hogy miért iktatunk be több lépcsőt a kiválasztás folyamatába. Ezek Kish (1995b) alapján a következők:

- A gazdaságban, társadalomban kialakuló „természetes” csoportok gyakran túl nagyok ahhoz, hogy teljes egészében megfigyeljük a bennük található egyedeket. Például, ha a településeket tekintjük elsődleges mintavételi egységeknek, akkor sok esetben ez túl sok sokasági elemet tartalmaz, és szükséges ezeknek a szűkítése a végső megfigyeléshez.
- Felmerülhet bennünk az a gondolat, hogy az előbbi problémát áthidalhatnánk azzal a megoldással is, hogy minden települést eleve kisebb csoportokra osztunk, és így máris alkalmazható a csoportos kiválasztás egy lépcsőben. Igen ám, de lehetséges, hogy minden elsődleges mintavételi egység csoportosítása túl nagy munkát igényelne, költséges és időigényes folyamat lenne. A többlépcsős módszerrel viszont elegendő csak az első körben kiválasztott PSU-k további csoportosítását elvégezni.
- A csoportos kiválasztás által okozott magasabb standard hibák gyakran csökkenthetők a több fázis alkalmazásával.

- Bizonyos megkérdezés-típusok esetében a csoportok teljes körű megfigyelése kivitelezhetetlen, vagy legalábbis nem hatékony. Ilyen lehet az, ha a PSU a háztartás, és a háztartás minden tagjával interjút kell készíteni, akkor valószínűleg nem megoldható, hogy az egyes családtagok ne hallják a többiek válaszait. Így viszont nem biztosítható, hogy független válaszok szülessenek.

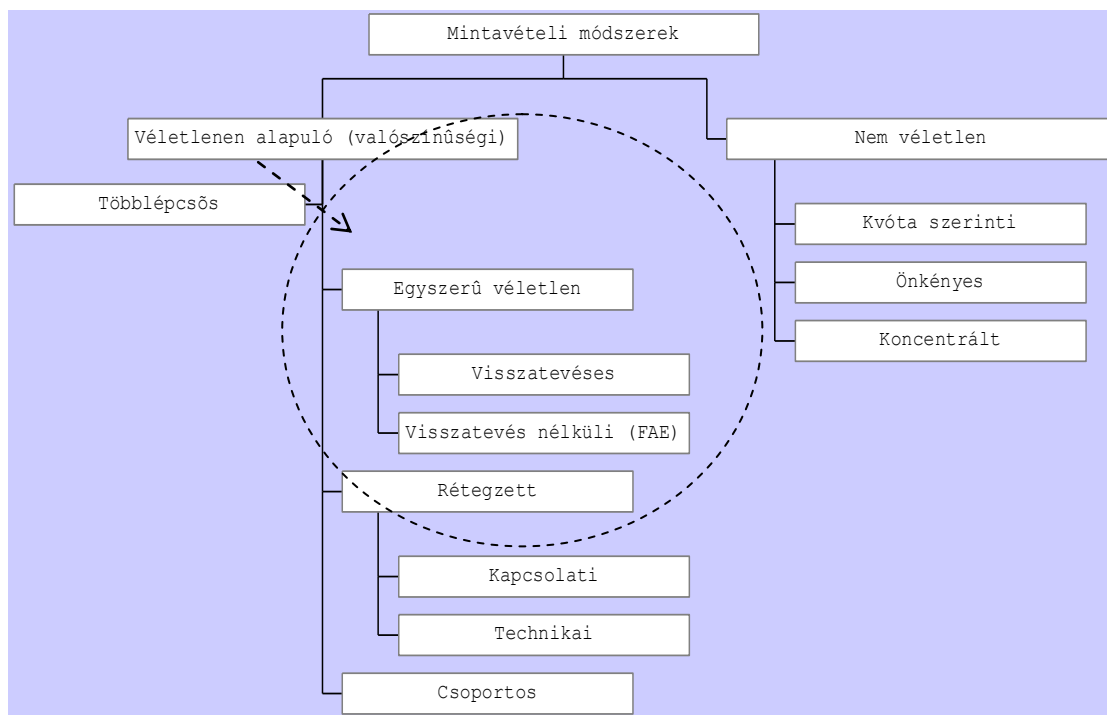
A többlépcsős módszerrel kiválasztott minta adataiból történő becslés esetében a standard hiba nagyságát a külső szórás mellett már a belső szórás is befolyásolja. Előbbit ugyanúgy tudjuk csökkenteni, mint a csoportos kiválasztás esetében. Utóbbi akkor lehet kisebb, ha növeljük a végső mintaelemek számát, illetve ha kapcsolati rétegzést végzünk az utolsó lépcsőben. Általánosságban annyit mondhatunk még, hogy a külső és a belső szórás arányainak vizsgálata vezethet egy optimális kiválasztási terv elkészítéséhez.

### 3.1.5. *Mintavételi módszerek csoportosítása*

A mintavételi módszerek – ezen belül is a véletlenül alapuló kiválasztások – egyértelmű csoportosítása valószínűleg megoldhatatlan feladat. A szakirodalomban talált csoportosítások természetesen valamilyen szempontrendszer mentén logikus felépítést mutatnak, és legfőbb elemeiben azonosak, mégis számomra egyik sem nyújtott megfelelő képet. Alapvetően a rétegzett és a többlépcsős mintavétel elhelyezésében láttam a problémát, ezért is próbálkoztam egy saját csoportosítási rendszer kialakításával, ami a további mondanivalómnak is megfelel (10. ábra). A rétegzett mintavételt két részre bontottam, aminek okaira már a korábbiakban kitértem. A többlépcsős mintavételt pedig mintegy a többi véletlenül alapuló mintavételi módszer valamilyen kombinációját jelölöm az összefogó halmazsal. Teszem ezt azért, mert önmagában a többlépcsős mintavétel nem létezik, az a többi mintavétel egymásra épülésével alakul ki. Ezért, véleményem szerint nem indokolt teljesen szeparáltan szerepeltetni.

A szisztematikus kiválasztást egyáltalán nem szerepeltetem az ábrában. Ennek oka az, hogy önmagában a szisztematikusság nem határozza meg a kiválasztással nyert minta jellegét. Ahogy már korábban említettem, a kiindulást jelentő lajstromunk rendezettsége határozza meg a szisztematikus kiválasztás eredményeinek tulajdonságait. Ezért tehát, véleményem szerint ez nem is egy mintavételi módszer, csak egy technikai, kivitelezési eszköz, mint például a véletlen szám generátorok felhasználása, vagy a sorsolás.

10. ÁBRA: A MINTAVÉTELI MÓDSZEREK CSOPORTOSÍTÁSA



Forrás: saját szerkesztés

### 3.1.6. Nem véletlen alapuló mintavételi módszerek

A 10. ábra is tartalmazza a nem véletlen alapuló mintavételi módszereket. Ezeknél a módszereknél fennállhat a torzítás veszélye, ugyanakkor körültekintő alkalmazásukkal hasznos információkhoz juthatunk. Ezeket csak röviden mutatom be (Hajdu et. al. 1994 alapján), mivel a dolgozatban kutatott téma feldolgozása során nem kerülnek felhasználásra.

**Kvóta szerinti kiválasztás:** Az összeíró kap egy összeírási körzetet, amelyen belül tetszőlegesen választhatja ki a megfigyelési egységeket. Ezt annyiban korlátozzák, hogy az egységek bizonyos, az összeírás témája szempontjából fontos jellemzői alapján arányokat határoznak meg (kvótákat állítanak fel), amelyeknek meg kell felelni. (Például 10 személyből legyen három 25 év alatti nő, kettő 25 év alatti férfi, négy 25 éves vagy idősebb férfi és egy 25 éves vagy idősebb nő.) Erre azért van szükség, mert ezáltal biztosítják az országos arányok megtartását, teljesülését a mintasokaságban. A minta összetételében, a vizsgált változó(k) vonatkozásában meg fog egyezni az alapsokasági megoszlásokkal.



**Önkényes (tudatos) kiválasztás:** Ekkor az összeíró saját megítélése alapján választja ki a sokasági egyedeket. Ez megfelelő megoldás lehet egy kisebb körű, általános benyomásokat mérő lekérdezés során (például egy áruházban a vásárlók pillanatnyi elégedettségének felmérésére), de nagyobb sokaságra, általánosabban nem vonhatóak le következtetések az eredményekből.

**Koncentrált kiválasztás:** Akkor választunk koncentráltan, céltudatosan elemeket, ha fennáll az a helyzet, hogy néhány kiemelt megfigyelési egység összeírása pontos képet ad a vizsgált jelenségről. Például az árak összeírása esetén elegendő lehet a legnagyobb, legtöbb üzlettel rendelkező kiskereskedelmi cégek megfigyelése, mert az árak alakulása ezen a koncentrált adatszolgáltatói körön is viszonylag pontosan tetten érhető.

### ***3.2. Mintavétel 2000 után a magyar mezőgazdaság statisztikában***

A mintavételi módszerek elméleti áttekintése után áttérek a mezőgazdaság statisztika területén alkalmazott gyakorlati kérdések bemutatására. A KSH-ban alkalmazott mintavételi módszereket 3 időszakra bontva tekintem át.<sup>17</sup>

#### ***3.2.1. 2000-2005***

A mezőgazdaság statisztikában a 10 évente zajló censusok közötti időszakban reprezentatív megfigyelésekkel biztosítják a folyamatos adatgyűjtést. A mintavételi módszer kialakítása, illetve a konkrét minta kiválasztása a teljes körű összeírások és a nagyobb lélegzetű gazdaságszerkezeti összeírások után zajlik le.

A 2000. évi ÁMÖ közel egymilliós gazdaságszámot mutatott a magyar mezőgazdaság háztartási oldalán. Nyilvánvalóan, az éves, évközi rendszeres összeírások lebonyolításához megfelelő minta kiválasztására volt szükség. A mintavétel módjának kialakítását a teljes körű

---

<sup>17</sup> Mivel a Hivatalnál töltött 5 év alatt (2002-2007) én voltam felelős a mintavételi és becslési módszerekért a főosztályon, ezért többször előfordul, hogy többes szám első személyben beszélek a munkáról, a kialakított módszerekről, hiszen azok kidolgozása a fenti időszakban elsősorban az én feladatom volt.

összeírás adatiból végzett próbaszámítások előzték meg. A statisztikai elméleti megfontolásokon túl pedig természetesen a pénzügyi korlát volt az, ami a minta nagyságát elsősorban meghatározta.

A 2000-2005 közötti időszakban csoportos, kétlépcsős, rétegzett mintavétel került kialakításra. A csoportos kiválasztást a korábbiakban már jelzett relatív „hatásosság” hiánya ellenére az összeírás költségvetési és szervezési oldala tette szükségessé. Az ÁMÖ során kialakított összeírási körzetek<sup>18</sup> jelentették az elsődleges mintavételi egységeket (14195 körzet) vagyis a csoportokat. Minden 8. összeírási körzet (a körzetek 12,5 százaléka) került kiválasztásra első lépcsőben. Ezeken a körzeteken belül a gazdaságokat 2 (illetve 3) réteg szerint csoportosították:

- **„A” réteg:** azok a gazdaságok, amelyek a következő küszöbértékek közül legalább az egyiket elérik: 5 szarvasmarha, 10 sertés, 26 juh, 26 liba, 5 ha szántó, 1 ha szőlő, 1 ha gyümölcsös, 100 tyúkféle, 100 kacsa, 100 pulyka, 25 méhcsalád („nagyobb” gazdaságok).
- **„B” réteg:** akik a gazdaságküszöböt elérték (lásd 1.4.1.), de a fenti, „A” réteghez tartozó küszöbértékeket nem („kisebb” gazdaságok).
- **(„C” réteg:** a gazdaságküszöb alatti háztartások, akiknél az ÁMÖ során mezőgazdasági tevékenységet regisztráltak, de ők csak 2001-ben szerepeltek az éves összeírás (OSAP1677) mintájában, és 2003 augusztusáig az állatállomány összeírásokban (OSAP1087).)

A rétegeket elválasztó állatállomány és földterület mértékek próbaszámítások alapján kerültek meghatározásra.

A „A” rétegből, a nagyobb gazdaságok közül a kiválasztott körzetekben minden gazdaság (100%) bekerült a mintába, míg a „B” rétegből, a kisebb gazdaságok köréből egyszerű véletlen alapján csak minden harmadik (33%). Az eltérő mintavételi arányra nemcsak azért

---

<sup>18</sup> Összeírási körzet: Akkora terület egy településen belül (lehet akár egy egész település is), amekkorát egy összeíró 3 hét alatt be tud járni a teljes körű összeírás céljából. Célszerűen utca, házszám lehatárolás alapján alakították ki a Megyei Igazgatóságok dolgozói a helyismeret birtokában. Maximum 150 címet tartalmaz egy körzet. Átlagosan 67 cím jut egy körzetre.

volt szükség, hogy a mintaelemszám megfelelő keretek között maradjon, hanem módszertani okai voltak. A „B” rétegbe tartozó kisebb gazdaságok esetében a fontosabb mutatók relatív szórásai lényegesen alacsonyabb szinten mozognak, hiszen ennek a rétegnek szigorú felső korlátai vannak. Ezért az egyes mutatók értékkészletének terjedelme kisebb, és pontosabban behatárolt. Itt tehát viszonylag homogén a gazdaságok köre, ezért kisebb mintavételi aránnyal is hatékonyabb becslésekre van lehetőség. Az „A” rétegben viszont a változók értékeinek határa csak a „csillagos ég”. A felső tartományban szinte minden mutató rendelkezik kiugró értékekkel (outlier), amelyek a szóródást extrém mértékben megnövelik. Ennek megfelelően a relatív szórások értéke is viszonylag magas. Ezért a homogenitás kritériumának ez a réteg kevésbé felel meg. A rétegen belüli teljes kiválasztás a heterogenitásból adódó nagyobb bizonytalanságot hivatott enyhíteni, mérsékelni.

A mintaválasztás során a megyék is egyfajta területi rétegzési szempontot jelentettek, tehát a körzetek egyszerű véletlen kiválasztására a megyék nagyságával arányosan került sor (technikai rétegzés), hiszen a körzeteket megyék szerint rendezték véletlen sorrendben a kiválasztás előtt.

Ez az „univerzális” minta szolgált alapul minden egyéni gazdaságokra vonatkozó összeírás (OSAP 1089 – állatállomány összeírás, OSAP 1651 – föld-, és vetésterület összeírás, OSAP 1677 – éves mezőgazdasági összeírás) során. Tehát nem volt külön állatállományra és külön földterületre vonatkozó mintavétel, hanem a fent említett szempontokat alkalmazták egy közös minta kialakítására. Az első állatállomány és földterület összeírások után összeírástechnikai szempontból annyi módosulás történt, hogy akik egyértelműen jelezték, hogy tartósan nem foglalkoznak növénytermesztéssel, illetve állattartással, azokat az összeíróknak nem kellett a következő alkalommal újra felkeresniük. Az ilyen jellegű információ úgynevezett címkódok segítségével került rögzítésre a KSH adatbázisában, amelyeket az összeírók a lajstromokon vezettek, és ezek adatai később még a kérdőívek előtt rögzítésre is kerültek. A minta összetételét ez természetesen nem befolyásolta. Az adatszolgáltatási teher és az összeírói munka csökkentése volt csak a cél. Ezek a gazdaságok 0-s adatokkal továbbra is szerepeltek a feldolgozásokban.

A konkrét mintaelemek kiválasztását a 2003-as Gazdaságszerkezeti Összeírás (GSZÖ) után újra elvégeztük (mintafriessítés), így egy gazdaság 3-4 évig szerepelt a mintában. Országosan mintegy 53 ezer gazdaságot tudtunk így megfigyelni, ami 5-6 százalékos mintavételi aránnak felelt meg akkor. Ez természetesen csak a mezőgazdasági teljesítmény

mintegy 50 százalékát produkáló egyéni gazdaságok körére érendő. A gazdasági szervezeteket emellett teljes körűen megfigyelték.

### 3.2.2. 2006-2010

A 2005-ös GSZÖ után újból időszerűvé és szükségessé vált a mintaelemek frissítése. A mintavétel elve alapvetően nem változott meg, de pénzügyi és egyéb nehézségek miatt a Hivatal vezetése a minta csökkentését kérte. Így az első lépcsőben csak minden 9. körzet került kiválasztásra (11,1 százalék). Második lépcsőben az „A” rétegre megmaradt a 100 százalékos kiválasztási arány. A további csökkentés a „B” réteget érintette. Itt a 4-es lépésközzre tértünk át (25 százalékos mintavételi arány). Országos szinten a mintaelemszám körülbelül 40 ezerre csökkent. A mintavételi arány a 2005-ös gazdaságszámot véve alapul tulajdonképpen nem változott, 5-6 százalék körül maradt, hiszen időközben az alapsokaság mérete is jelentősen lecsökkent. Ugyanakkor itt kell hangsúlyoznom azt, hogy a regiszter szintjén jelentkező alapsokaság, és a tényleges gazdaságszám itt már jelentős különbségeket mutatott. Ahogy távolodunk a census évtől, úgy válik el egyre inkább egymástól a gazdaságszám és a regisztrált alapsokaság. Ez egyrészt a mezőgazdaság még mindig gyors ütemű változásából fakad. Másrészt az ÁMÖ után nincsen lehetőség olyan információgyűjtésre, ami minden gazdaságra kiterjedne. Ebből adódik, hogy a megszűnt, mezőgazdasági tevékenységet már nem végző gazdaságokat csak abban az esetben tudjuk regiszterinformáció szintjén rögzíteni, ha konkrétan az adott gazdaság bekerült a rendszeres összeírások vagy a GSZÖ mintájába. Mindez egyben azt is jelenti, hogy az újonnan létrejött gazdaságokat sem tudjuk megfigyelni, rólunk sincsen információ két census között. Sajnos nincsen más forrás a gazdaságok létezésének folyamatos frissítésére az egyéni gazdaságok körében. Ez pedig azt jelenti, hogy a gazdaságok közel 2/3-áról nem áll rendelkezésre 2000-nél frissebb információ. Tehát gyakorlatilag a mintavételi keret bizonyos idő eltelte után mindig nagyobb, mint a tényleges gazdaságszám. Ebből az következik, hogy a minta aránya a regisztrált keretsokasághoz képest végül is lecsökkent. Nyilvánvaló tehát, hogy az elvek változatlansága és a mintaelemszám ilyen típusú csökkentése az összeírás mintavételi hibáit megnövelte volna. A hibaszámokat azonban szeretnénk volna a korábbi szintnek megfelelően megtartani. Javaslatomra az a kiegészítő megoldás született, hogy témánként (állatállomány és földterület szempontjából) kijelölésre került egy úgynevezett kiemelt egyéni gazdaságokat

tartalmazó kör. Ez a kis csoport postai úton teljesítette az adatszolgáltatást a gazdasági szervezetekhez hasonlóan.

Az **állatállomány** esetében a következő küszöbértékeket állítottuk fel: 100 szarvasmarha, 300 sertés, 50 ló, 750 juh, 100 kecske, 20000 tyúk, 5000 liba, 5000 kacsa, 5000 pulyka, 250 gyöngyös, 500 házinyúl, 500 vágógalamb, 400 méhcsalád (a mutatók között vagylagos összefüggés van). Így összesen 807 kiemelt gazdaságot kaptunk.

A **földterület** szempontjából az alábbi küszöbértékeket használtuk: legalább 250 ha szántó vagy 17 hektár szőlő vagy 22 hektár gyümölcsös. Így 796 kiemelt gazdaságot találtunk. Ebből 35 az állatállomány szempontjából is kiemelt volt.

Az év végi, minden témát átfogó összeírás (OSAP1677) esetében a 2 féle kiemelt kör uniója képezte a teljes körűen megfigyelt egyéni gazdaságok körét.

A kiemelés alapjául szolgáló mutatók kiválasztása során több szempontot is figyelembe vettünk:

- Jogszabályi előírások: melyek azok az állat, illetve növényfajok, amelyekre vonatkozóan előírják a maximálisan teljesítendő hibahatárokat.
- Lefedettségi arány: mi az a küszöbérték, ami alapján az egyes mutatók értékének megfelelő lefedettségét biztosítani tudjuk, és ezzel a hibaszint javulása elegendő mértékű.
- Kiválasztott adatszolgáltatók száma: A kiemelt kör kiválasztása során az adatszolgáltatók számossága nem lehetett túl nagy pénzügyi és összeírás-technikai okokból sem.

A végül kiválasztott „legnagyobb” gazdaságokat teljes körűen megfigyeltük. Ez más szóhasználatnál élve azt is jelenti, hogy az „outliereket”<sup>19</sup>, a kiemelkedően nagy értékekkel (földterület, állatállomány vagy mindkettő terén) rendelkező gazdaságokat 100 százalékban összeírjuk (a későbbiekben is tárgyalt „take all” elv a magas értékekkel rendelkező

---

<sup>19</sup> Ezeket a nagyobb gazdaságokat azért tekinthetjük „kiugró” elemeknek, mert a teljes sokaságot figyelembe véve, felfelé „kilógnak” a méretük miatt, ráadásul számosságuk is viszonylag kicsi.

gazdaságok körében). Ennek nyilvánvaló előnye az, hogy mind az állattartás, mint a növénytermesztés szempontjából az adatok egy jelentős hányadát teljes körűen, mintavételi szempontból 0 százalékos hibával tudjuk begyűjteni. Ezáltal a „maradék” mintavételes megfigyelés statisztikai hibája kisebb súllyal esik latba. Ugyanakkor azt is el kell ismerni, hogy a gyakorlatban nem számít minden szempontból sikertörténetnek a kiemelt kör alkalmazása. A postai adatgyűjtés, illetve a helyenként helytelen név és címadatok miatt viszonylag nagy arányú ebben a körben a nemválaszolás mértéke, ami eddig, az összeírók által lebonyolított adatgyűjtésekben szinte ismeretlen jelenség volt. A kiemelt körben 2008 decemberében 2,5 százalékos, 2009 decemberében pedig 3,2 százalékos nemválaszolási arány jelentkezett. Ha a „normál” összeírási gyakorlatot nézzük, ez nem tekinthető különösen magas aránynak, ugyanakkor ha számításba vesszük azt is, hogy itt a jelentős földterülettel, illetve állatállománnyal rendelkező gazdaságokról van szó, akkor ez a pár százalék már nagyobb súllyal esik latba.

### 3.2.3. 2010 után

A 2010-es census után új minta kiválasztására került sor. A mintaválasztás módszertana a korábbi időszakokhoz képest némileg módosult. Az egyéni gazdaságok kijelölése a mintába a 2010. évi Általános Mezőgazdasági Összeírás alapján történt. Az ország összes mezőgazdasági (összeírási) körzetéből (kb. 14 000) 407 került kijelölésre (a körzetek kb. 3 százaléka). Mivel további kiválasztási lépcső beiktatására nem került sor, ez egy csoportos mintavételi technika. A kijelölt körzeteket tehát teljes körűen írták össze az összeírók. A mintába az összes egyéni gazdaság 5,3 százaléka, mintegy 30 ezer egyéni gazdaság került. Annyiban ugyanakkor speciális az összeírás, hogy a számlálóbiztosok nem csak a 2010-es ÁMÖ alapján egyéni gazdaságnak minősülő háztartásokat keresték fel, hanem „teljes körzetbejárást” végeztek, vagyis házról-házra járva az esetleg új egyéni gazdaságokat is megtalálták. Így nem csak a lajstromon szereplő gazdaságokról, hanem az „újjakról” is kérdőívet töltenek ki az összeírók. Fontos kiemelni azt is, hogy a körzeteken belül tehát, a korábbiakkal ellentétben nem alkalmaznak további rétegzést.

A fentiekén túl megmaradt a korábbiakban bevált úgynevezett kiemelt egyéni gazdaságokat tartalmazó kör megfigyelése is. Az alkalmazott küszöbértékek módosultak, legtöbb esetben csökkentek. Állatállomány esetében a következők szerint: 150 szarvasmarha,

300 sertés, 25 ló, 500 juh, 40 kecske, 20 000 tyúk, 5000 lúd, 10000 kacska, 5000 pulyka, 120 gyöngyös, 300 házinyúl, 200 vágógalamb, 300 méhcsalád, 50 strucc, 10 emu, 200 prémes állat. A földterülethez használt küszöbértékek: 250 hektár szántó, 17 hektár szőlő, 20 hektár gyümölcsös, 100 hektár erdő, 150 hektár gyeplő. A legalább egy feltétel megléte esetén kiválasztott „kiemelt” gazdaságok száma közel 3300. A „take all” elv alkalmazása a nagy gazdaságok körében tehát továbbra is segíti a jobb minőségű adatok előállítását. Ez a teljes körűen megfigyelt rész postai úton teljesíti az adatszolgáltatást, így összeírás-technikai szempontból továbbra sem jelent gondot az, hogy ezek a gazdaságok földrajzilag viszonylag szétszórtan helyezkednek el. Ugyanakkor a magyarországi adatszolgáltatási tradíciók és kultúra helyzete miatt a postai adatgyűjtés még a viszonylag nagy gazdaságok esetében is akadózva halad. Viszonylag sok a nemválaszolás, ami sajnos befolyásolja az adatok minőségét. Mégis az a véleményem, hogy valahol el kell kezdeni az adatszolgáltatók „bevonását”, és önkéntesre ösztönözni őket, hogy idővel megszokják a postai vagy esetleg az online válaszadást, és ezzel hosszabb távon csökkenhessenek az összeírás költségei, és egyidejűleg javuljon az adatok minősége is.

Összességében azt gondolom, hogy kedvező lehet az új módszernek azon vonása, hogy minden háztartást felkeresnek a körzeten belül. Ezzel az a cél, hogy ne csak a megszűnő gazdaságokat mérjék fel, hanem az időközben gazdaságküszöb fölé kerülő háztartásokat is regisztrálni tudják, és ezzel az alapsokasági elemszámra folyamatosan egy pontosabb becslést lehessen kalkulálni. A rétegzés mellőzése viszont egy hatékonyságot javító eszközt hagy kihasználatlanul. Az utólagos rétegzés is lehet további opció a becslések pontosságának javításához. Ugyanakkor továbbra is az összeírás szervezésének, lebonyolításának szempontjai az elsődleges meghatározó tényezők a minta kialakítása során. Az elvárt pontosság elérése ezért véleményem szerint veszélyben van, elsősorban a juhállomány szempontjából.

A mintavételi gyakorlat áttekintésével azt is szerettem volna érzékeltetni, hogy a gyakorlati munka során az elméleti lehetőségeket meglehetősen leszűkített formában lehet csak felhasználni. Az összeírások pénzügyi háttere, az összeírók beszervezhetősége, az egy adatszolgáltatóra fordítható összeírási költség mind-mind olyan tényezők, amelyek viszonylag szűk kényszerpályát jelölnek ki a mintavételi terv kialakítására. Nem az elméleti lehetőségek, a kreatív megoldások kiaknázásának terepe ez. Sem idő, sem pénz, sem elegendő szakmai támogatás nincsen meg a hatékonyabb módszerek kidolgozásához. Éppen ezért szeretném

kutatóként feldolgozni és szimulálni a szélesebb körű elméleti lehetőségeket, természetesen a későbbi gyakorlati alkalmazás reményében.

### ***3.3. A statisztikai adatok minősége***

Minden statisztikai felvétel célja jó minőségű adatok létrehozása. Ez egy triviálisnak ható kijelentés, ugyanakkor korántsem egyszerű feladat a minőségnek, illetve annak értékelésének a definiálása, meghatározása. Mivel az EU-s előírások a mintavételi hiba maximális nagyságának meghatározása révén kívánják kijelölni egy-egy összeírás minőségének a kereteit, ezért foglalkoznunk kell a mintavételi hiba sajátosságaival is. De azt is látni kell, hogy a hiba (legyen az mintavételi vagy nem mintavételi) a statisztikai adatok minőségének csak egy dimenziója. Egy fejezetrész erejéig szükségesnek tartom áttekinteni a statisztikai adatok minőségének definícióját, összetevőit. Mégpedig azért, mert a dolgozatban csak a mintavételi hibára helyeződik a hangsúly, de a kép akkor teljes, ha a minőség kérdésének többi aspektusát is látjuk.

A statisztikai minőség fogalmát meghatározni nem könnyű feladat. Időben változó jelenségről van szó, ami függ a felhasználói igényektől és a statisztikai termék jellegétől is. Kezdetben az adatok pontosságával azonosították, de a fogalom értelmezése folyamatosan bővült. Az Európai Unió statisztikai gyakorlatában nagyon sok helyen találkozunk a minőség meghatározásával. 2003 végére született meg az úgynevezett minőségjelentés véglegesítése. Ennek a munkának az egyik eredménye a statisztikai folyamat minőségének tisztázása, a minőségjelentés formájának meghatározása mellett a termékminőség egységes dimenzióinak a kialakítása és értelmezése volt (Szép – Vigh 2004). Ez alapján tehát 6 dimenziót kell felsorolni:

- Relevancia (relevance)
- Pontosság (accuracy)
- Időszerűség (timeliness)
- Hozzáférhetőség és érthetőség (accessibility, clarity)
- Összehasonlíthatóság (comparability)
- Koherencia (coherence)



Folyamatosan kísérleteznek azzal, hogy a dimenziók jellemzőit egy mutatóba, egyetlen minőségindexbe tömörítsék, de eddig nem született minden szempontot kielégítő megoldás. Nem célom részletesen bemutatni a dimenziókat, de érzékeltetni szerettem volna, hogy a mintavételi hiba önmagában nem is teljes dimenzió, hanem „csak” a pontosságnak az egyik része a nem mintavételi hibák mellett.

Az összeírások pontossága, legyenek azok mintavételes, vagy teljes körű felvételek, természetesen nem csak abból tevődik össze, hogy a sokaság milyen részét mérjük fel. A minőséget ennél lényegesen több tényező befolyásolja. Mivel a pontosságnak két fő összetevője van ezért a mintavételi hiba mellett fontos kitérni a nem mintavételi hibák jellemzőire, előfordulási típusaira is. A statisztikai adatok, következtetések pontosságát elemi és aggregált szinten egyaránt a „valós” és a megfigyelt érték közötti különbséggé kellene értelmezni. Mivel azonban a valós értékek nem állnak rendelkezésre (éppen ezért van szükség adatgyűjtésre), ezért meg kell elégednünk azzal, hogy a hibák különböző előfordulási típusait próbáljuk meg mérni, számszerűsíteni, és az eredmények fényében értelmezzük az adatok pontosságát.

A jelenségek megfigyeléséből nyert információ természetesen csak bizonyos korlátok között vezethet helytálló következtetésekre. A korlátok értelmezéséhez, belátáshoz meg kell vizsgálni a választott módszerekkel együtt járó sajátosságokat, amelyek hatással vannak a statisztikai következtetések érvényességére.

A megfigyelés lehet teljes körű (cenzusok) vagy részleges, mintavételen alapuló. A statisztikai munka során keletkező hibák egyik része (nem mintavételi hibák) érintheti a teljes körű összeírásokat is. A mintavételi hiba pedig csak a részsokaság összeírásával megvalósuló felvételek sajátja. És fontos hangsúlyozni, hogy itt nincsen feltételes mód. Ez természetszerűleg minden mintavételes megfigyelésnek sajátossága. Ebben az esetben tehát a „hiba” szó használata nem jelenti azt, hogy valamit rosszul csináltunk volna, egyszerűen a módszer egyik jellemzője. Annak csak a mértéke változó. Azonban ezt is tovább árnyalhatjuk. A mintavételi hibának is két forrása lehet. Az egyik a véletlen mintavételi hiba, ami abból adódik, hogy nem a teljes sokaságot figyeljük meg. A másik lehet a torzítás, ami a becslőfüggvény torzított voltából következik. A mintavételi hibát nem tudjuk kiküszöbölni, viszont a torzítás elkerülhető torzítatlan becslőfüggvények alkalmazásával (torzítatlan = a becslőfüggvény várható értékében megegyezik a becslni kívánt változóval). A nem véletlen mintavételi módokkal szemben az egyszerű véletlen mintavételi módnak éppen az az előnye,

hogy az alapsokaságra vonatkozó becslések hibahatára magából a mintából egzakt módszerekkel, matematikai számítások útján meghatározható, illetve a mintanagyság növelésével „tetszőlegesen” csökkenthető.

A nem mintavételi hibák gyakorlatilag minden adatgyűjtés során felmerül(het)nek. Főbb típusaik:

1. a mintavételi keret nem mindig fedi le a vizsgálni kívánt sokaságot, nincsenek megfelelő információink az alapsokaságról
2. a kiválasztott minta egyedei nem mindig hajlandók a közreműködésre (nemválaszolás),
3. az egyes válaszok nem hibátlanok (mérési hibák), ide tartoznak az úgynevezett szisztematikus torzítások is (például az éves jövedelmet általában alul, míg a kiadásokat hajlamosak vagyunk felülbecsülni a kérdőívek kérdéseire válaszolva), mérési hibák (Párniczky 1966.; Mundruczó 1979)
4. feldolgozási hibák, elírások,
5. (vannak, akik (Szép – Vigh 2004) a modell-feltételezések hibáit is ide sorolják).

A legtöbb mintavételi módszer esetében feltétlenül szükséges, hogy rendelkezésre álljon egy alkalmas mintavételi keret, ha úgy tetszik az alapsokaság megfelelő lajstroma, listája, amiből a mintaválasztás megtörténhet. Itt máris felmerül az első problémaforrás: a mintavételi keret teljessége (minősége), vagyis a lefedettség. A tényleges alapsokaságtól való eltérés mindkét irányba hibaforrást jelent. Amennyiben nincsen minden elem a felsorolásban, akkor un. undercoverage-ról beszélhetünk. Ekkor tehát bizonyos elemek kimaradtak a lajstromból, ezért nem kerülhetnek be a mintába sem. Hogyha tudjuk, hogy vannak kimaradó elemek, és külső információk alapján van ismeretünk ezek elemszámáról is, akkor a hiba némileg korrigálható a megfelelő alapsokasági elemszám (N) felhasználásával a becslési eljárás során. Amennyiben nem tudjuk, hogy vannak kimaradó elemek, illetve nem tudjuk ezek számosságát, akkor a felszorzással is csak a kisebb lefedettséget tudjuk elérni. Az overcoverage is problémát jelent, bár véleményem szerint kevésbé kockázatos. Ebben az esetben ugyanis olyan elemek is szerepelnek a mintavételi keretben, akik nem tartoznak oda. A nem oda tartozás oka kettős lehet: egyrészt már megszűntek, csak a regiszter még nem tud erről; másrészt létező egységek ugyan, de nem felelnek meg az összeírás célsokaságának (mert például nem végeznek mezőgazdasági tevékenységet). Ha tehát minden, a célsokaságnak megfelelő egyed szerepel a mintavételi keretben, és azon felül vannak oda nem

illőek is, akkor ez véleményem szerint nem okoz a pontosság kérdésében problémát. Hiszen ekkor a véletlenszerű mintaválasztás során bekerülnek a nem odaillő elemek is, a kérdésekre adott válaszaikból pedig kiderül, hogy nem részei a célsokaságnak. Gyakorlatilag 0 értékeket hoznak az adatok közé, és a felszorzás során így a célsokaságra jellemző tényleges értékeket kaphatjuk vissza. Igaz viszont az is, hogy a valós, információt hordozó mintaelemek száma kevesebb lesz, mint a tervezett mintanagyság, ha azonban a mintanagyságot a nagyobb célsokaságot figyelembe véve alakítottuk ki, akkor az sem okoz különösebb problémát. Az egyes változókra jellemző szórásokat módosíthatják ezek a nullás értékek, ha nem kezeljük külön a célsokaságon túli csoportot. Azonban arra is van mód, hogy a mintavétel során nyert információk alapján a célsokaság elemszámát ( $N$ ) is módosítsuk (megbecsüljük), és ekkor kihagyhatóak a célcsoporton kívüli kérdőíveken kapott válaszok.

A gyakorlatban egyik típus sem szerepel külön-külön. Általában, tetézve a problémát, minden irányban vannak hibái a mintavételi keretnek. Tehát tartalmaz célsokaságon kívüli elemeket, ugyanakkor az alapsokasághoz tartozó elemek egy része pedig hiányzik. Ilyenkor a legnehezebb a helyzet. A mintavételi keret tehát alapvetően befolyásolja a kiválasztható mintát, annak reprezentativitását.

Ha sikerül is kiválasztani egy megfelelő mintasokaságot, akkor is számolni kell azzal, hogy a ténylegesen megfigyelt egyedek száma ennél biztosan alacsonyabb lesz.

Nem válaszolásról akkor beszélhetünk, ha bizonyos egyedekről nem sikerül a kívánt információt részben vagy egészben megszerezni. A nem válaszolás jelensége, amennyiben a kérdőív kitöltése teljes egészében megghiúsult, jól látható, számszerűsíthető kategória. Nem válaszolási arányt is szoktak számítani, ami a kitöltetlen kérdőívek számát viszonyítja a kijelölt minta nagyságához. A másik oldalról megközelítve a visszaérkezési arány mutatója is használatban van. Ez természetesen a nem válaszolási arány komplementere (visszaérkezési arány és a nem válaszolási arány együttesen kiadják a 100 százalékot). Ekkor tehát van képünk a megghiúsulások nagyságrendjére vonatkozóan, legalábbis az elemszámok vonatkozásában. Ugyanakkor, problémát jelent az, hogy a nem válaszoló sokasági elemről egyáltalán nincsen, vagy csak nagyon régi adatok állnak rendelkezésre. Vagyis a kérdőíven szereplő változók hiányzó értékeire csak hozzávetőleges információval rendelkezünk. A megvalósult minta átsúlyozásával, kalibrálásával a nem válaszolásnak ez az esete viszonylag hatékonyan kezelhető.

Nem ennyire egyszerű a helyzet akkor, ha részleges nem válaszolásról van szó. Ez azt jelenti, hogy az adatszolgáltató összességében teljesítette az adatszolgáltatási kötelezettségét, de bizonyos kérdésekre elfelejtett válaszolni, vagy szándékosan elhallgatta az adatokat. Az ilyen hiányokat a legnehezebb kimutatni és mérni. A mezőgazdaság statisztika jellegéből adódóan alkalmas arra, hogy a kérdőívben szereplő táblázatok közötti összefüggések segítségével a részleges nem válaszolást azonosítani lehessen, de így is lehetnek megbújó hiányosságok.

A nem válaszolások kiküszöbölésének legjobb módszere a személyes megkérdezés, de ez nem valósítható meg minden esetben, mert nagyon költségigényes megoldás. A kérdésfeltevés, az ellenőrzési szempontok, a táblák, kérdések közötti navigáció segítése mind-mind hatékonyabbá tehetik az adatok begyűjtését. A másik lehetőség a nem válaszolás érzékelése esetén annak okainak feltárása, és azok orvoslása, megszüntetése. Ez azonban meglehetősen nehéz feladat. (A témában lásd: Szilágyi 2011.)

A mérési hibák jelentik a következő problémaforrást. Ez azt jelenti, hogy a válaszokban pontatlan adatok szerepelnek. Akár azért mert a válaszadó nem tudja pontosan megválaszolni a kérdést (nem emlékszik például a felhasznált műtrágya típusára, mennyiségére), vagy rossz mértékegységet használ, vagy egyszerűen elírja a választ. Az adatgyűjtések során szerzett tapasztalataim alapján úgy látom, hogy a mérési hiba okozza a legkisebb károkat, előfordulása a legritkább, és kiküszöbölése a legjobban megoldható. A kérdőívekben kialakított összefüggések, ellenőrző kérdések, összegző sorok felhasználásával nagyon jól elkerülhetők ezek a problémák. A pontos, tömör, közérthető kitöltési útmutatók is sokat segítenek a megfelelő minőségű adatok megszerzésében.

Az adatok rögzítése során is keletkezhetnek hibák. Az előző esethez hasonlóan elütésekről, elírásokról, kihagyott értékekről lehet szó. A megfelelő adatrögzítő programok beépített ellenőrzési szempontjai a legtöbb esetben automatikusan figyelmeztetnek a hibákra, de természetesen az elütések minden típusa ellen nem lehet védekezni. Utólagos ellenőrzési lehetőségek is vannak. Ezekkel a módszerekkel elsősorban a kiugróan magas vagy alacsony értékeket lehet kiszűrni, és orvosolni a hibás rögzítést. A rejtve maradt hibák nagyságára viszont sajnos nincsen megfelelő mérési módszer. Hiszen, ha tudnánk, hol vannak a hibák, akkor a javításukra is sor kerülhetne, nem kellene megmérni a nagyságukat.

Láthatjuk tehát, hogy a mintavételi hiba – mint a pontosság kritériumrendszerének egyik alkotóeleme – mellett a nem mintavételi hibák szerepe is fontos, és nem elhanyagolható. A nem mintavételi hibák kezelése során nagyon fontos, hogy figyeljünk arra, mely hibák előfordulása tekinthető véletlennek, és melyek azok, amelyek valamilyen szisztematikus torzítás eredményei. A legnagyobb gondot az utóbbiak szokták okozni.

A jogszabályi előírások a minőség dimenziói kapcsán csak olyasmit tudnak előírni és elvárni, amelyek egyértelműen mérhetőek. A nem mintavételi hibák nem feltétlenül tartoznak ebbe a kategóriába. Ezekről inkább a minőségjelentések keretében kell nyilatkozni, sokszor csak közelítő értékekkel, az összeírás körülményeinek szöveges megfogalmazásával. A gyakorlatban tehát ezért kerül előtérbe a mintavételi hiba nagysága akkor, amikor pontosságról beszélünk. Ez egyértelműen számszerűsíthető, és éppen ezért ellenőrizhető is.

A dolgozatban a későbbiekben bemutatott kísérleti számítások (szimulációk), elméleti fejtegetések és módszertani újítások kidolgozásának éppen az a célja, hogy a mintavételes összeírásokból származó becsült adatok minőségét javítani lehessen az eredmények felhasználásával. A nem mintavételi hibákkal nem foglalkoznak a számítások, hiszen azok nehezen számszerűsíthetőek, esetlegesek, és összeírásról-összeírásra változhatnak. Ugyanakkor fontos kiemelni, hogy a nem mintavételi hibákkal a gyakorlatban foglalkozni kell, azok potenciális nagyságára vonatkozóan mindig kell információt gyűjteni.

## **4. A szimulációs számítások háttere, az adatállomány jellemzői**

A kutatómunka lehetőségét gyakorlati szempontból a KSH-ban szerzett tapasztalatok, és az adatbázisok ismerete adta. Technikailag a KSH korábbi elnökének engedélyével ideiglenes felhasználóvá váltam a KSH Pécsi Regionális Igazgatóságán. Itt egy számítógép állt rendelkezésemre – szinte korlátlanul – a lekérdezések futtatására. A mezőgazdasági összeírásokból származó alapadatokat tartalmazó adatbázisokhoz, adattáblákhoz kaptam hozzáférést. Természetesen annyi megkötéssel, hogy egyedi adatokat nem láthatok, és az általános adatvédelmi szabályokat betartom. A lekérdezéseket SQL plus-ban végeztem. A kutatáshoz szükséges alapadatok lekérdezéseit ennek megfelelően az SQL nyelv felhasználásával magamnak kellett megírni. A másik fontos feltétel az adatbázisok ismerete, illetve a KSH-ban használatos névkonvenciókban, szabványokban való jártasság volt. Ezek nélkül nehezen lehetne eligazodni a nómenklatúrák, mutatók elnevezései, struktúrái között. A számításokhoz körülbelül 10 adatbázis táblát kellett felhasználnom, összekapcsolnom.

### ***4.1. A kísérleti adatbázis***

A dolgozat célja az alapsokaság jellegzetességeinek megismerésén és bemutatásán túl a különböző mintavételi technikák eredményeinek értékelése szimulációs eljárás segítségével. Ehhez a kísérleti jellegű mintakiválasztásokat, számításokat természetesen célszerű volna egy aktuális alapsokaságon végrehajtani. Ha azonban ilyen naprakész teljes körű adatbázis mindenkor rendelkezésre állna, akkor nyilván nem is volna szükség mintavételre.

Az utolsó elérhető teljes körű összeírás 2000-ben zajlott. Természetesen az ott feljegyzett közel 1 millió gazdaság adatai adatbázisban elérhetőek, de a mezőgazdaság rendkívül gyors ütemű változása (az egyéni gazdaságok száma 2007-re 35 százalékkal esett vissza) ezt a csaknem 10 éves nyilvántartást – éppen a vizsgálódások célja miatt – használhatatlanná tette. A legfrissebb, nagyobb lélegzetvételű adatgyűjtést a 2007-es GSZÖ jelenti. (A 2003-as szerkezeti összeírás körülbelül háromszor nagyobb mintán került végrehajtásra, de az azóta eltelt 9 év azokat az adatokat is „elavulttá” teszi a vizsgálatok szempontjából.) A 2010. évben végrehajtott census adatai a dolgozat befejezésének idejére álltak rendelkezésre. Éppen ezért

az olvasó csak az 5. fejezet végén (5.4. fejezet) találkozik a legfrissebb census adataiból készített „próbaszámítások” eredményeivel.

A 2007-es GSZÖ mintájának alapsokasághoz viszonyított aránya körülbelül 18 százalék. Ennek kiválasztásához csoportos kiválasztást alkalmazott a Mezőgazdasági főosztály. A régiókon belül véletlen számok segítségével sorba rendezett összeírási körzetekből minden 6. került a mintába. Ez önmagában 16,7 százalékos kiválasztási arányt jelentene, de ehhez jönnek hozzá a kiemelt (extrém nagy) gazdaságok, illetve a körzetek eltérő nagysága is növelte a mintavételi arányt. Az összeírás során teljes körzetbejárás történt. Ez azt jelenti, hogy nem csak a korábbi információk alapján a lajstromra került egyéni gazdaságokat keresték fel a számlálóbiztosok, hanem minden, mezőgazdasági termelés szempontjából potenciális háztartást meglátogattak. Így lehetőség volt a megszűnéseken túl a mérleg másik oldaláról, a keletkező egyéni gazdaságokról is adatokat begyűjteni a kiválasztott körzetekben. Kísérleti számításaim során ezt a GSZÖ 2007 során összeírt, mintegy 111 ezer gazdaságot fogom alapnak tekinteni. A minta nagysága és a kiválasztás módja együttesen biztosítják azt, hogy ez a mintasokaság az alapsokasághoz nagyon hasonló eloszlásokat, megoszlásokat követ az egyes mutatók szempontjából országos és regionális szinten egyaránt. A kísérleti számításokhoz a fent bemutatott GSZÖ 2007-es mintaállományt, mint alapsokaságot tekintem, és a továbbiakban alapsokaságként fogok hivatkozni rá, az alapsokasági jelölésrendszernek megfelelően. (Szilágyi Roland, korábban már hivatkozott disszertációjában is hasonlóan járt el a Háztartási költségvetési felvételt vizsgálva.)

A szükséges mintaelemszámok meghatározásához szükség lesz az alapsokasági értékösszegek megengedhető hibanagyságára. Ez mindig az adott szemlélet szerint kialakított átlagos értékek nagyságának a függvényében alakul ki.

A számítások, szimulációk végső célja egy optimális mintavételi terv kialakítása. Optimális alatt azt értem, hogy a jelenleg alkalmazott 30-40 ezer közötti mintaelemszámot lehetőleg ne haladja meg, és ennek a leghatékonyabb elosztását találjam meg, ami a 3 kiemelt állatkategória szempontjából a legkisebb mintavételi hibát eredményezi, illetve megfelel az Európai Unió fentebb említett jogszabályban rögzített követelményeinek. A tudományos kutatás köntösébe bújva az összeírhatóság és az összeírás költségigényessége szempontrendszerrel másodlagosan kezelem. A Hivatali munka során a szervezeti és pénzügyi korlátok jelentősen behatárolják a választási lehetőségeket. Az elméleti megfontolások szélesebb spektrumának vizsgálatára sokszor nem jut elég idő, erőforrás. Ettől függetlenül

természetesen az elért eredmények alapján azt remélem, hogy a KSH összeírási rendszerébe is be tudnak majd épülni a kidolgozott megoldások valamilyen kompromisszumos formában.

#### ***4.2. Az állattartás speciális jellemzői***

A jelenlegi mezőgazdasági statisztikai rendszerben mintavételi szempontból nem különítik el az állattartó gazdaságokat, ahogy ez a mintavételi módszerből ki is derült már korábban. Ugyanakkor létezik kétféle megközelítés is, ami a gazdaságokat „profiljuk” szerint csoportosítja.

Az egyik a KSH által használt hármas csoportosítási elv, amelyben növénytermesztő, állattartó és vegyes gazdaságokat különítenek el. Sok táblázatot találunk ebben a bontásban a GSZÖ-k kiadványaiban. A csoportosítás definíciója viszonylag egyszerű. A gazdaságküszöb definíciójában felsorolt mutatókat és azok nagyságát veszik alapul. Az a gazdaság, aki csak az állatállomány szempontjából éri el a küszöböt, állattartó, aki csak a termőterülete alapján minősül gazdaságnak, növénytermesztő, aki pedig mindkét szempont szerint átlépi a küszöbértékeket, vegyes gazdaságnak minősül. Ez igazából nem jellemzi a specializáció jellegét, hiszen, ha egy gazdaságnak jelentős mennyiségű termőterülete van, mégis tart mellette néhány állatot, akkor már vegyes gazdaságnak minősül. Az egyes ágazatok súlyát ez a fajta csoportosítás nem igazán használja ki.

Az Eurostat által kidolgozott gazdaságtipológiai rendszer<sup>20</sup> ennél lényegesen finomabban hangolt csoportosítást jelent. A gazdaságszerkezeti összeírások és a censusok eredményeiből kell ezt a fajta csoportosítást elvégezniük a tagállamoknak a részletesen meghatározott szempontrendszer és döntési fa felhasználásával. A korábbi tipológiában 8 főcsoport (illetve van egy 9. is, ami inkább technikai jellegű, és a nem besorolható gazdaságokat tartalmazza, akik magyar viszonylatban többnyire az erdőgazdálkodással foglalkozó egységek) és további 3 szinten több mint 50 alcsoport került meghatározásra. A gazdaságok profilját 1/3-os, 2/3-os arányhatárok felállításával lehet meghatározni. Vagyis az egyes típusokba való besoroláskor

---

<sup>20</sup> A Bizottság 85/377/EGK határozata (1985. június 7.) a mezőgazdasági üzemek közösségi tipológiájának kialakításáról.



nemcsak egy alacsony küszöbértéket használnak fel, hanem a gazdaság összes termelésén, tevékenységén keresztül a súlypontokat keresik, és ez alapján kerülnek a megfelelő csoportba. Az egységes mérőszám az SFH érték. A gyakorlati besorolást egy döntési fa rendszerének megfelelően kell elképzelni. Első körben meghatározzák, hogy a 9 főcsoporton belül melyikbe tartozik a gazdaság (melyik tevékenysége képviseli a legnagyobb részt). Ezután következik az alábontás a további szintekbe. Ahhoz azonban, hogy mondjuk szántóföldi növénytermelő főcsoportba kerüljön egy gazdaság, nem elég nagyobb részt ezzel foglalkoznia, hanem legalább a teljesítményének 2/3 részét kell ide tartozóan elérnie. Ez az elv minden további szintre is érvényes.

Ezt a rendszert némileg megváltoztatta a Bizottság 1242/2008/EK rendelete a mezőgazdasági gazdaságok közösségi tipológiájának létrehozásáról. Továbbra is 8 (illetve 9) főcsoport van, de ez alatt már csak 2 további szinttel. A rendszer lényegét tekintve nem történt jelentős módosítás, inkább a csoportokon belüli átrendeződésekről van szó. Ami fontos változás még, hogy az SFH értékről át kell majd térni egy úgynevezett standard termelési értékre, ami az új tipológia kialakításának alapját is fogja jelenteni. (Hozzáteszem, az időbeli összehasonlíthatóság feltételének teljesülése csorbulni látszik a változtatások bevezetésével.)

A létező csoportosítások véleményem szerint nem alkalmasak mintavételi célokra. Az első túlságosan egyszerű, mégsem „tisztá” csoportokat hoz létre. A tipológia pedig túl bonyolult, még akkor is, ha csak a 8 főcsoportot vesszük figyelembe. A továbbiakban én három szemléletben fogom vizsgálni a gazdaságokat:

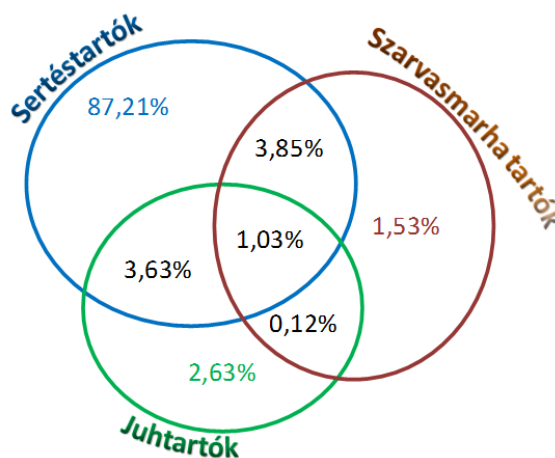
- összes gazdaság (profilra való tekintet nélkül mindenki)
- állattartó gazdaságok (bármilyen tenyészállatot tartó gazdaság)
- állatfajt tartó gazdaságok (a 3 vizsgált állatfaj szerint külön-külön itt csak azokat szerepeltetem, akik az adott állatfajjal rendelkeznek).

A továbbiakban az állattartás, állattartók jellemzőit fogom számba venni és megvizsgálni a mintavételi szempontokat szem előtt tartva.

Alapvető probléma a mezőgazdaság statisztikában, hogy egy-egy gazdaság egyszerre többféle mezőgazdasági tevékenységgel foglalkozik, ráadásul eltérő volumenben, nagyságrendben. Egyszerre többféle állatot tart, de 1-2 fajra koncentráltan, vagyis míg van nagy számban szarvasmarha állomány, mellette csak 1-2 sertést, illetve néhány juhot találunk.

Ezért fontosnak tartom megvizsgálni az egyéni gazdaságok jellemzőit a többcélúság szempontjából. Hangsúlyozom, hogy a most következő adatok csakis „technikai” célokat szolgálnak. Az állattartó gazdaságokra átlagosan az a jellemző, hogy 1,86 állatfajjal foglalkoznak egyszerre. (Ebbe beleértendő az összes tenyészállat-fajta.) Vagyis az esetek többségében csak egy-kétféle állatfajt tenyésztenek. Az átlagos értéken túl részleteiben is érdemes megvizsgálni, hogyan alakulnak a 3 kiemelt fajt tartó gazdaságok esetében az átfedések. A 11. ábra halmazok segítségével szemlélteti a gazdaságok megoszlását a különböző csoportokban (az összes gazdaság, aki a három állatfaj közül legalább az egyik tartja jelenti a 100 százalékot). Egyértelműen látható, hogy a sertéstartók számossága és aránya a legmagasabb. A gazdaságok 8,6 százaléka az, aki legalább kétféle állat tenyésztésével foglalkozik egyszerre. A másik oldalról ez azt jelenti, hogy 91,4 százalékuk a három faj közül egyre „specializálódott”. Az arányeltolódás nagyon erőteljes, ugyanakkor a bonyodalmat mintavételi szempontból mégiscsak az a kevésnek tűnő 8,6 százalék okozza.

11. ÁBRA: SERTÉST, JUHOT, ILLETVE SZARVASMARHÁT TARTÓ GAZDASÁGOK

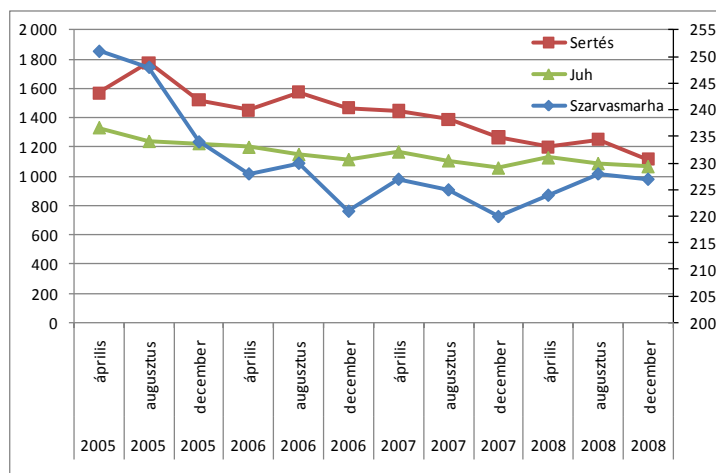


*Forrás: KSH adatbázis alapján, saját szerkesztés*

Az is fontos szempont a mintavételi terv kialakítása során, hogy az állattartó gazdaságok mennyire változékonyak időben. Vagyis mennyire jellemző az, hogy profilt váltanak (átternek sertéstartásról juhtartásra például), vagy bővítik a profiljukat (a baromfi és sertéstartás mellé szarvasmarhákcal is elkezdnek foglalkozni), illetve felhagynak vagy éppen belekezdnek az állattartásba. Mintavételi szempontból az állandóság lenne a legkedvezőbb tulajdonság. Hiszen minél stabilabbak a tartás szerint kialakítható rétegek, annál pontosabban kalkulálhatóak a várható hibaszámok az összeírások előtt. A rendelkezésre álló információk

alapján azonban ezt a fajta változékonyságot a legnehezebb meghatározni. Teljes körű információ csak a 2000-es ÁMÖ adatbázisában található, a követő években csak mintavételes megfigyelésekre került sor. Vagyis az egyes gazdaságok viselkedését, változását csak néhány esetben lehet nyomon követni (ha adott gazdaság mintaelem is volt valamelyik, esetleg több évben is). Mivel a 2003-as GSZÖ-ben is viszonylag nagy mintán hajtották végre az összeírást (körülbelül 30 százalékos volt a mintavételi arány), és az 1 EUME<sup>21</sup> feletti gazdaságok teljes körűen szerepeltek a mintában, így még a 2003-as év az, ami a legjobb támpontot adhatja. A becsült eredmények azt mutatják, hogy 2000 és 2007 között az állandó állattartók aránya 70-90 százalék között mozog. Ez véleményem szerint az ágazat viszonylagos stabilitását támasztja alá. Csoportosítási, rétegzési szempontból ez jó hír. Az előzetes információk alapján kialakított homogén részsokaságok fő jellemzői csak minimálisan térnek el a ténylegesen összeírtaktól. Ugyanakkor, mivel a stabilitás nem tökéletes, mindig külön figyelmet kell szentelni a változások vizsgálatára is.

12. ÁBRA: AZ ÁLLATÁLLOMÁNY SZEZONÁLIS INGADOZÁSA AZ EGYÉNI GAZDASÁGOKBAN  
(EZER DB)<sup>22</sup>



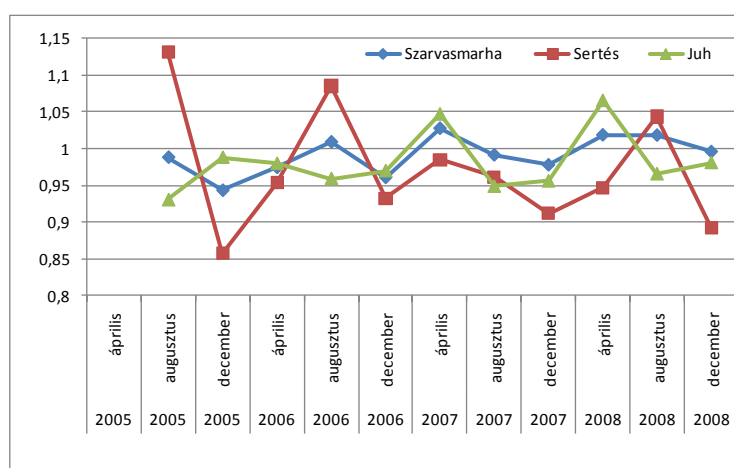
Forrás: [www.ksh.hu](http://www.ksh.hu) (Agrár idősorok és censzusok), saját szerkesztés

<sup>21</sup> EUME: Európai Méretegység (European Size Unit – ESU). 1 EUME 1200€ SFH értéknek felel meg. Minden gazdaságot besorolnak az SFH érték alapján az EUME szerinti méretkategóriákba.

<sup>22</sup> A bal oldali függőleges tengely a sertés- és juhállomány nagyságát, míg a jobb oldali a szarvasmarhaállomány nagyságát mutatja.

Mindezek mellett még egy fontos szempontot szem előtt kell tartania a statisztikusnak. Az állattartás szezonális jelenség (lásd 12. ábra). Vagyis az év során két alkalommal sorra kerülő összeírások az egyes állatfajok szintjén más-más nagyságot képviselnek. Ez természetes velejárója az állattartás jellegének, ugyanakkor az állattartó gazdaságok meghatározása során erre is figyelemmel kell lenni. A korábbi évek összeírásaiból ez a szezonális jelleg még inkább szembetűnő, hiszen 2008-ig évi 3 alkalommal gyűjtötték be az adatokat. A szezonális ingadozás nagyságának szemléltetésére lásd a 13. ábrát. (A sertésciklusokról lásd 2.1 fejezet.) A sertés és a szarvasmarha – az állomány nagyság függvényében – augusztusban van „csúcson”, és decemberben jellemző a legalacsonyabb állatszám. A juhok esetében az áprilisban jelentkezik az éven belüli legmagasabb érték.

13. ÁBRA: AZ ÁLLATÁLLOMÁNY SZEZONÁLIS VÁLTOZÁSAI AZ EGYÉNI GAZDASÁGOKBAN  
(ELŐZŐ IDŐSZAK = 100%)



Forrás: [www.ksh.hu](http://www.ksh.hu) (Agrár idősorok és censzusok), saját szerkesztés

### 4.3. Az aszimmetria mérése a mezőgazdaságban

Már az állatállomány adatok részletes bemutatása, elemzése előtt is érzékelhető, és a gyakorlati tapasztalatok alapján is látható, hogy mindhárom állatfaj esetében aszimmetrikus, más szóval ferde eloszlással lesz dolgunk. Ez alatt azt értem, hogy a gazdaságok az általuk tartott állatok száma alapján aszimmetrikus eloszlással rendelkeznek, és a továbbiakban ezt fogom egyszerűen az állatfajok eloszlásának nevezni. Mivel azt gondolom, hogy az adatsorok aszimmetrikus jellege alapvetően befolyásolja a mintavételi módszerek hatékonyságát és alkalmazhatóságát, ezért célszerű még a részletes leíró statisztikai elemzés előtt összefoglalni az aszimmetria fogalmával, mérésével és minősítésével kapcsolatos tudnivalókat.

Az aszimmetriát felfoghatjuk úgy, mint a szimmetria hiánya. Egy eloszlás akkor tekinthető szimmetrikusnak, ha a mediántól jobbra és balra is ugyanolyan képet mutat. Más szavakkal ez azt is jelenti, hogy a helyzeti középértékek (a módusz és a medián) valamint a számtani átlag egy pontba esnek. Ennek a szimmetriának a hiánya jelenti azt, hogy az eloszlás ferde, vagyis egyik irányba elnyúlik a gyakorisági poligon vagy a sűrűségfüggvény széle. A ferdeségnek tehát iránya is van. Lehet jobb, illetve bal oldali. A nemzetközi szakirodalomban nem teljesen egységes az irányok elnevezése. Pintér-Rappai (2007) alapján a dolgozatban a sűrűségfüggvény megnyúlt része adja meg az aszimmetria oldalát. Vagyis amennyiben az átlag nagyobb, mint a helyzeti középértékek, akkor jobb oldali, ellenkező esetben bal oldali aszimmetriáról beszélhetünk. A legtöbb mutató esetében a pozitív értékek a jobb oldalra, míg a negatív értékek a bal oldali ferdeségre utalnak.

Az aszimmetria számszerűsítésére számos mutatót dolgoztak ki matematikusok, illetve a statisztika művelői. A legegyszerűbb mutatók a számtani átlag és a helyzeti középértékek kapcsolatán alapulnak. A Pearson-féle „A” mutató a módusz és a számtani átlag távolságát méri a szórás arányában:

$$A = \frac{\bar{x} - Mo}{\sigma}$$

ahol:

Mo – módusz

$\bar{x}$  – számtani átlag

$\sigma$  – szórás

A mutató szimmetria esetén a nulla értéket veszi fel, mivel ekkor az átlag és a módusz megegyezik. Előjelével jelzi az aszimmetria irányát, viszont abszolút értékben nézve nincsen felső korlátja. Vagyis az aszimmetria erősségét illetően nemigen tudunk mit mondani. Ugyanilyen tulajdonságokkal bír az  $A_2$ -vel jelölt mutató, ami a medián és az átlag távolságát használja fel a ferdeség mérésében.

$$A_2 = \frac{3(\bar{x} - Me)}{\sigma}$$

ahol:

Me – medián

Többféle mutatót alakítottak ki a kvantilisek felhasználásával is. Az első, F-mutató elnevezésű a kvartiliseket használja fel, illetve a szimmetriának azt a jellemzőjét, miszerint az alsó és a felső kvartilis egyenlő távolságra helyezkedik el a mediántól.

$$F = \frac{(Q_3 - Me) - (Me - Q_1)}{(Q_3 - Me) + (Me - Q_1)}$$

ahol:

$Q_1$  – alsó kvartilis

$Q_3$  – felső kvartilis

Az F-mutató lehetséges értékei – a képletből adódóan – zárt intervallumban helyezkednek el. Minimum -1, legfeljebb pedig +1 lehet. Ugyanezt a mutatót  $S_4$ -gyel is jelölik (Elsayed 2010-es cikkében B-vel jelöli), és más formában is felírható<sup>23</sup>:

$$S_4 = \frac{Q_1 - 2Me + Q_3}{Q_3 - Q_1}$$

Ennek egy másik változatában az oktiliseket használják:

$$S'_4 = \frac{O_1 - 2Me + O_7}{O_7 - O_1}$$

ahol:

$O_1$  – első oktilis

---

<sup>23</sup> Innentől kezdve Mandrekar et.al. (1995) cikkében bemutatott jelöléseket használom fel.

O<sub>7</sub> – hetedik oktilis

A ferdeségi együttható a centrális momentumokat használja fel az aszimmetria méréséhez.

$$S_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{\sqrt{\left( \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \right)^3}} = \frac{m_3}{(m_2)^{3/2}}$$

ahol:

Ugyanezen mutató mintából becsült értéke<sup>24</sup>:

$$S_1' = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{\sigma} \right)^3$$

Ennek a mutatónak a lényege, hogy az egyedi értékek távolságát vizsgálja meg az átlaghoz képest, a szórással elosztva. Vagyis standardizált értékeket használ fel. Alapvetően a normális eloszláshoz méri a ferdeséget. Akkor mutat szimmetriát, ha nulla értéket vesz fel. Előjele az előbbieknél megfelelően az aszimmetria irányát mutatja. Hasonlóan az A és A<sub>2</sub> mutatókhoz, az S<sub>1</sub> sem korlátos, ami nehezíti az értelmezhetőséget (már ami az aszimmetria erősségét illeti).

A ferdeség L-mutatója<sup>25</sup> az L-momentumokon (sorba rendezett sokaság lineáris kombinációi) alapul. Ez azt jelenti, hogy sorba rendezett sokaságból számszerűsíthető a következő formula:

<sup>24</sup> Az SPSS programcsomag és a Microsoft Excel is ezt a változatot használja az aszimmetria számszerűsítésére.

<sup>25</sup> Royston 1992-es cikkében ezt a mutatót  $\tau_3$ -mal jelöli, Randles et. al. 1980-as tanulmányában b<sub>1</sub>-et használják.

$$S_2 = \frac{l_3}{l_2}$$

ahol:

$$l_2 = 2w_2 - \bar{x}$$

$$l_3 = 6w_3 - 6w_2 + \bar{x}$$

$$w_2 = \frac{\sum_{i=2}^n (i-1)x_{(i)}}{n(n-1)}$$

$$w_3 = \frac{\sum_{i=3}^n (i-1)(i-2)x_{(i)}}{n(n-1)(n-2)}$$

$x_i$  – a korábbiaktól eltérően a nagyság szerint rendezett sokaság  $i$ -dik elemét jelöli

A mutató értéke  $-1$  és  $+1$  közé esik.  $0$  és  $-1$  között bal oldali,  $0$  és  $+1$  között jobb oldali aszimmetriát jelez. Természetesen, abszolút értékben minél magasabb számot kapunk, annál erőteljesebb az aszimmetria. Mivel a mutató egy normált értéket ad eredményül, hüvelykujj-szabályként azt fogalmazom meg – a későbbi elemzések elősegítése érdekében –, hogy abszolút értékben  $0,3$  alatt gyenge,  $0,3-0,7$  között közepes és  $0,7$  fölött erős aszimmetriáról beszélhetünk.

Hosking (1990) definiálta egy módosított, vagy származtatott verzióját is a ferdeség  $L$ -mutatójának:

$$S_2' = \frac{1 + S_2}{1 - S_2}$$

Ez a változat szemléletesebben értelmezhető. A mutató  $0$ -tól végtelenig vehet fel értékeket.  $S_2 = 1$  esetben mutat szimmetriát. Amennyiben  $1$ -nél kisebb, akkor bal oldali, ha pedig  $1$ -nél nagyobb, akkor jobb oldali aszimmetriát mutat. A mutató értéke pedig megmutatja, hogy milyen arányban áll az eloszlás felső széle az alsóval. Mivel a mutató egy normált értékből indul ki, ezért azonos mintanagyság mellett nem látom akadályát annak,



hogy a mutató nagyságrendje összehasonlítható legyen. Így a különböző állatfajok esetén ezzel a lehetőséggel élni is fogok.<sup>26</sup>

Egy másik megközelítés (Mandrekar et.al. 1995) szintén kvartiliseket használ, mégpedig percentiliseket, mintabeli információra alapozva. Ennek a módszernek az a lényege, hogy képezzük a következő különbségeket:

$$Me - P_{L_i}, \text{ illetve } P_{U_{100-i}} - Me \quad (i = 0, 1, 2, \dots, 49)$$

ahol:

$P_L$  – alsó percentilisek

$P_U$  – felső percentilisek

Szimmetria esetén az alsó és felső percentilis párokból képzett különbségeknek egyenlőnek kell lennie:

$$Me - P_{L_i} = P_{U_{100-i}} - Me, \text{ vagyis } Me' = \frac{P_{L_i} + P_{U_{100-i}}}{2}$$

Így kapunk 50 becslést ( $Me'$ ) a mediánra. Ezután 95%-os megbízhatóság mellett az átlaghoz tartozó becslés standard hibáját (SH) véve felső korlátnak, kiszámítjuk a következő kifejezés értékét:

$$p = \frac{\sum_{i=0}^{49} p_i}{50} \text{ ahol } p_i = 1, \text{ ha } |Me' - Me| \leq SH, \text{ egyébként } p_i = 0 \quad (i = 0, 1, 2, \dots, 49)$$

Ezzel a kifejezéssel p-re egy olyan értéket kapunk (0-1 intervallumban), ami becslést ad arra az arányra, hogy a mediánra vonatkozó becslés a standard hiba által meghatározott

---

<sup>26</sup> A mutatók sorából az  $S_3$  hiányzik, ami valójában nem is igazán mutató, hanem egy, a szimmetriára vonatkozó hipotézisellenőrzés próbafüggvénye. Mivel azonban a dolgozat szempontjából ez nem lényeges, részletesen nem tárgyalom. A leírása megtalálható Mandrekar et.al. (1995) cikkének 7-8. oldalán.

pontossággal hányszor egyezik meg a „tényleges” mediánnal. A cikkben hüvelykujj-szabályként nevesítenek is három kategóriát a mutató értékének értelmezéséhez:

- ha  $p \geq 0,9$ , akkor az empirikus eloszlásunk szimmetrikusnak tekinthető
- ha  $0,8 \leq p < 0,9$ , akkor valószínűleg van néhány kiugró érték az adatsorunkban
- ha  $p < 0,8$ , akkor az empirikus eloszlásunk nem szimmetrikus, és további vizsgálatok szükségesek

Mielőtt még részletesen megvizsgálom a három állatfaj jellemzőit, nézzük meg csak az aszimmetriát mérő mutatók alakulását, használhatóságát különböző sokasági szemléletben. Ezeket az értékeket tartalmazza az 5. táblázat.

5. TÁBLÁZAT: AZ ASZIMMETRIA ALAKULÁSA ÁLLATFAJONKÉNT, KÜLÖNBÖZŐ MUTATÓKKAL

Állatfaj	Mutató	Sokasági szemlélet		
		Fajt tartó	Állattartó	Összes
Sertés	A	0,280	0,283	0,238
	A <sub>2</sub>	0,593	0,550	0,713
	S <sub>1</sub>	17,909	21,307	24,561
	S <sub>2</sub>	0,689	0,645	0,708
	S <sub>2</sub> '	5,431	4,634	5,849
	S <sub>4</sub>	0,333	0,000	1,000
	S <sub>4</sub> '	0,667	0,600	1,000
	p	0,480	0,420	0,080
Szarvasmarha	A	0,530	0,120	0,102
	A <sub>2</sub>	1,071	0,361	0,306
	S <sub>1</sub>	3,292	15,111	17,824
	S <sub>2</sub>	0,540	0,958	0,953
	S <sub>2</sub> '	3,348	46,619	41,553
	S <sub>4</sub>	0,429	*	*
	S <sub>4</sub> '	0,625	*	*
	p	0,220	0,920	0,940
Juh	A	0,390	0,084	0,071
	A <sub>2</sub>	1,010	0,251	0,213
	S <sub>1</sub>	4,510	20,909	24,659
	S <sub>2</sub>	0,670	0,942	0,950
	S <sub>2</sub> '	5,061	33,483	39,000
	S <sub>4</sub>	0,579	*	*
	S <sub>4</sub> '	0,857	*	*
	p	0,040	0,920	0,940

\*a mutató nevezőjében nullás érték szerepel

Forrás: KSH adatbázis alapján, saját számítás

A mezőgazdaság jellegzetességei között megfogalmaztuk már korábban is, hogy egy-egy változót tekintve a teljes sokaság jelentős része nullás értéket vesz fel, mert adott állatfaj tartásával egyáltalán nem foglalkozik (ez a jelenség persze a földterületnél is tetten érhető). Ez azt jelenti, hogy a módusz, a legjellemzőbb érték nem ritkán a nulla lesz (lásd 4.2 fejezet). Erre azt is szokás mondani, hogy az aszimmetria szempontjából szélsőséges eloszlásokkal, úgynevezett J alakú eloszlásokkal van dolgunk, hiszen a módusz a legkisebb értékkel esik egybe. (Hajdu et. al. 1994) A **sok nullás érték**<sup>27</sup> esetünkben még azt is eredményezi, hogy a medián is ezt az értéket veszi fel, illetve a medián alatti összes kvantilis is. Ez mindenképpen speciális helyzetet jelent az aszimmetria mérésének szempontjából, és az **S<sub>4</sub>**, illetve **S<sub>4</sub>'** **mutatókat alkalmatlanná teszi a ferdeség mérésére.**

A Pearson-féle A mutató esetében a példánk kapcsán szintén érdekes következtetésre juthatunk. Mivel tehát sok esetben a módusz értéke nulla, az A mutató képlete egyszerűsödik, illetve átalakítható:

$$A = \frac{\bar{x}}{\sigma} = \frac{1}{V}$$

ahol:

V – relatív szórás

Vagyis az aszimmetria ilyen speciális esetben a relatív szórás reciprokának tekinthető a Pearson-féle A mutatónál. Ez azt jelentené, hogy minél magasabb a relatív szórás, annál alacsonyabb az aszimmetria mértéke. Azonban a szórás növekedése önmagában egyértelműen nem befolyásolja a szimmetriát vagy aszimmetriát, tehát ez biztosan nem jó irány. A végső következtetése a gondolatmenetnek az, hogy **a Pearson-féle A mutató egyszerűen nem alkalmas az aszimmetria mérésére ilyen speciális (sok nullás értéket tartalmazó) sokaságok esetében.** Kifejezetten félrevezető eredmények születhetnek, ha ezt használjuk fel.

---

<sup>27</sup> Sokat foglalkoznak más szempontból, főleg regressziós modellek becslése kapcsán hasonló sokaságokkal a szakirodalomban. Ezt a jelenséget angolul „zero-inflated” vagy „excess zeros” kifejezésekkel illetik. Néhány kapcsolódó cikk a témában: Heilbron 1994, Mullahy 1997, Ridout et. al. 1998, Staub-Winkelmann 2011.

Könnyen belátható, hogy az  $A_2$  mutatónál is hasonló probléma merül fel. Itt a medián lesz az, ami sok esetben nulla. Ezzel az  $A_2$  mutató képletét is felírhatjuk a következő formában:

$$A_2 = 3 \frac{\bar{x}}{\sigma} = 3 \frac{1}{V}$$

Vagyis itt is a relatív szórás növekedésével csökkenni fog az aszimmetria, ami nem helytálló következtetés, ezért az  **$A_2$  mutatót sem tudjuk esetünkben felhasználni.**

A számítások elvégzése nélkül is azt „érezzük”, hogy – a sertések kivételével – a sokasági szemlélet bővítésével, a jobb oldali aszimmetria erősödik, hiszen az amúgy is ferde eloszlás elejére jön még meglehetősen nagy számosságú nullás érték.

A táblázatban p-vel jelölt, a korábban bemutatott, percentilisek vizsgálatán alapuló mutatóval szintén problémák vannak a sok, állatot nem tartó gazdaság miatt. Ennél a módszernél, mutatónál az jelzi a nagyobb mértékű aszimmetriát, ha minél kisebb a kiszámított érték. Az állatfaj tartásával foglalkozó sokaság esetében úgy tűnik, jól is „működik” ez a számítás, hiszen a legkevésbé a sertések, leginkább a juhok mutatnak aszimmetriát. A másik két sokasági szemlélet azonban mind a juhok, mint a szarvasmarhák esetében egyértelműen növeli a ferdeséget, de egészen mást mutatnak a p-vel jelzett értékek. Mindkét állatfajnál azt tapasztaljuk, hogy a sokasági szemlélet bővítésével 0,9 fölé emelkedik a mutató értéke, ami a fenti bemutatás alapján azt jelentené, hogy az empirikus eloszlásunk szimmetrikusnak tekinthető, és ez nyilván nem igaz. Ebben az esetben is a sok állattartással nem foglalkozó gazdaság adja a probléma okát. Nem csak a viszonyítási alap, a medián veszi fel a nullás értéket, hanem a percentilisek jó része is. Ezért adódhat az a helyzet, hogy sokszor adnak jó becslést az egymáshoz képest szimmetrikusan elhelyezkedő alsó és felső percentilisek átlagai a mediánra. Ettől azonban nem lesz szimmetrikus az eloszlás. **A percentilisek elemzésén alapuló aszimmetria vizsgálat, illetve mérés tehát esetünkben szintén nem használható.** Érdekes ugyanakkor, hogyha a sertések példáját vesszük elő, és nézzük meg alaposabban. Ugyanez a mutató „jó irányba változik”, vagyis a sokaság bővülésével növekszik, és mindhárom esetben kimutatja az aszimmetriát. Ennek az az oka, hogy a sertéstartók száma (több, mint 51 ezer gazdaság) lényegesen nagyobb, mint a másik két faj esetében (a szarvasmarhánál és juhoknál a tartók száma nem éri el a 4 ezret sem). Ezért a sertést nem tartó gazdaságok aránya az állattartókat, illetve az összes gazdaságot magában foglaló sokaság esetében sem lesz annyira magas, mint a másik két esetben. Ezért – ugyan a teljes

sokaságot figyelembe véve a medián már 0 lesz – a percentilisek medián fölötti része már jellemzően nem nulla.

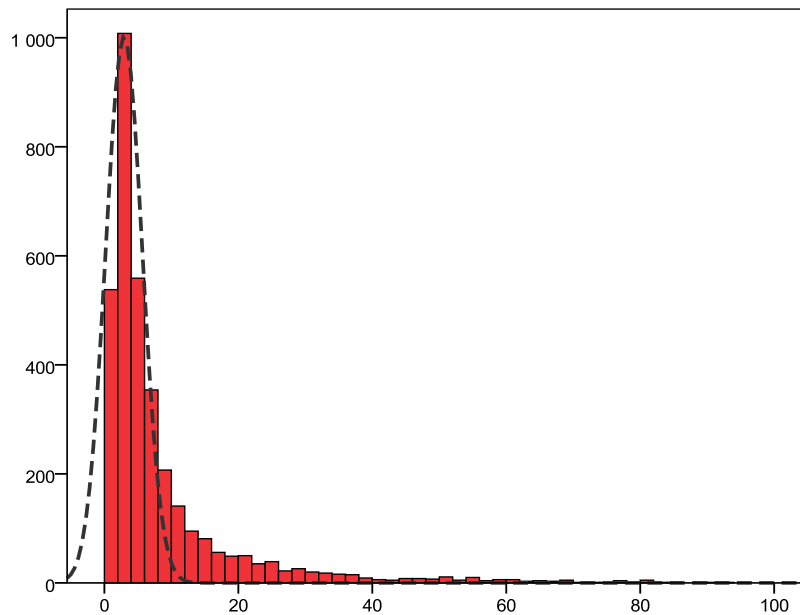
Az  $S_1$  mutató viszonylag jól elterjedt, sok szoftver is ezt használja. Ugyanakkor ennek a mutatónak is vannak **gyengeségei**. Royston (1992) foglalja össze, hogy melyek a mutató hátrányai. Egyrészt **nehezen értelmezhető** a kapott eredmény, legalábbis, ami az aszimmetria erősségét illeti. Másrészt a mutató nagyon **érzékeny** az eloszlás széleiben végbemenő változásokra. A hivatkozott cikkben odáig megy a szerző, hogy azt javasolja, a szoftverekben is váltani kell, a sokkal jobban értelmezhető, és kevésbé érzékeny  $S_2$ , illetve  $S_2'$  mutatók használatára. Ezek talán egyetlen hátránya a jelenős számítási igény. Ez azonban a számítástechnikának és a szoftvereknek köszönhetően nem okozhat gondot a mai világban.

Az aszimmetria mérési sajátosságainak a bemutatása végén tehát arra a következtetésre jutottam, hogy az állatállomány leíró elemzésére a ferdeség méréséhez az  $S_2$  és az  $S_2'$  **mutatókat fogom felhasználni**, mint az empirikus adatokhoz mindenféle szempontból leginkább igazodó indikátorokat. A mutatók konkrét értelmezését ennek megfelelően a következő fejezet tartalmazza a többi leíró statisztikai elemzéssel együtt.

#### ***4.4. Leíró elemzés állatfajonként***

A különböző állatfajok eloszlásának vizsgálatára és szemléltetésére a legalkalmasabb eszköz a hisztogram lenne. Ennek azonban gátat szab az a tény, hogy az eloszlások meglehetősen aszimmetrikus jelleget mutatnak. Mindhárom állatfajt tartó gazdaságokra az a jellemző, hogy 1-2 állattal rendelkeznek nagy tömegben, és szinte az összes többi nagyságkategória gyakorisága elenyésző ezekhez képest. Ezért a hisztogramon annyit látnánk, hogy van egy magas kezdő oszlop, majd az x tengellyel szinte teljesen egybeolvadva a többi. A leginkább „érzékelhető” képet a szarvasmarha hisztogramja adja (lásd 14. ábra). Ez a hisztogram az illusztrálhatóság kedvéért, nem tartalmazza azokat a gazdaságokat, akiknek nincsen szarvasmarhájuk. (A sertés és a juh gyakorisági adataiból szerkesztett hisztogram a IV. számú mellékletben található.) A szaggatott vonal az átlagnak és szórásnak megfelelő normális eloszlást mutatja. Jól látható, hogy ettől nagymértékben eltérnek az empirikus gyakoriságok.

14. ÁBRA: A SZARVASMARHA ÁLLOMÁNY HISZTOGRAMJA



*Forrás: KSH adatbázis alapján, saját szerkesztés*

Az alapsokasági jellemzők átfogó bemutatásához tehát további vizsgálati eszközökre van szükség.

A nagyságrendek érzékeltetéséhez osztályközös gyakorisági sorokat is készítettem. Az osztályközös gyakorisági sorokba rendezés elveinek megtartása nehézségekbe ütközött az extrémítások miatt. Az egyenlő hosszúságú osztályközök alkalmazása nagymértékben „elkendőzne” az állatállomány eloszlások jellegzetességeit, így eltérő hosszúságú csoportokat kellett kialakítani. Ebben az esetben viszont fennáll annak a veszélye, hogy a gyakoriságokat közvetlenül összehasonlítjuk egymással, és helytelen következtetéseket vonunk le. Ennek elkerülésére két további statisztikai eszközt is segítségül hívtam. A leíró statisztika mutatóinak számbavétele megint egy lépéssel közelebb visz az adatsorok jellegzetességeinek megragadásához. Végül pedig Lorenz-görbék segítségével szemléltetem az állatállományok koncentrációjának alakulását.

A szarvasmarha állomány eloszlása jobb oldali aszimmetriát mutat. A ferdeség mutatóiból azt olvashatjuk le, hogy csak a szarvasmarhát tartókat tekintve is közepes erősségű

aszimmetriát látunk. Az állattartó, illetve a teljes sokaság pedig megközelíti az  $S_2$  mutató elméleti maximumát. Az eloszlás jobb oldala több, mint 40-szerese a bal oldalnak. Érdekes továbbá, hogy a legintenzívebb aszimmetria nem a teljes sokaságnál, hanem az állattartóknál jelentkezik (bár a különbség nem nagy). Ennek az lehet az oka, hogy bár előbbi esetében az egy gazdaságra jutó állománynagyság valamelyest csökken, a képzeletbeli hisztogramunk elejére bekerül a sok állat nélküli gazdaság, némileg kiegyensúlyozva a két oldalt.

A legjellemzőbb az, hogy a gazdaságoknál 2 szarvasmarha található, de jellemző a mindössze 1 állatból álló, illetve a 3-6 közötti állománynagyság is. A 11 alatti állománnyal rendelkezők az összes szarvasmarhát tartók 80 százalékát teszik ki, ugyanakkor mindössze az állomány 28 százalékát birtokolják. Érdekessége még az eloszlásnak, hogy az 50 feletti állatszám esetében is viszonylag nagy a változatosság. A tartók száma egyre csökken, ugyanakkor az állománynagyság jelentős. (Lásd: 6. táblázat.)

6. TÁBLÁZAT: SZARVASMARHA-ÁLLOMÁNY OSZTÁLYKÖZÖS GYAKORISÁGI SORA

Szarvasmarha-állomány nagyságkategóriái	Gyakoriság (gazdaságok száma)
1-2	1129
3-6	1188
7-15	666
16-60	428
61-	36
Összesen	3 447

*Forrás: KSH adatbázis alapján, saját számítás*

7. TÁBLÁZAT: A SZARVASMARHA-ÁLLOMÁNY LEÍRÓ STATISZTIKÁJA KÜLÖNBÖZŐ SOKASÁGOK SZERINT

Mutató	Szarvasmarhát tartó gazdaságok köre	Állattartó gazdaságok köre	Összes gazdaság
Gazdaságok száma	3 447	79 837	110 949
Átlag	8,13	0,35	0,25
Medián	4	0	0
Módusz	2	0	0
Minimum	1	0	0
Maximum	93	93	93
Terjedelem	92	93	93
Szórás	11,57	2,92	2,48
Ferdeség ( $S_2$ )	0,540	0,958	0,953
Ferdeség ( $S_2'$ )	3,348	46,619	41,553
Értékösszeg	28 032		

*Forrás: KSH adatbázis alapján, saját számítás*

A maximális állatállomány viszonylag alacsony, nem éri el a 100-at. Ebből arra következtethetünk, hogy mintavételi szempontból a szarvasmarha esetében lesz a legkönnyebb a dolgunk. A viszonylag kicsi terjedelem miatt a sokaság változatossága is meghatározott keretek között marad, ezért könnyebben reprezentálható egy mintasokasággal. (Lásd: 7. táblázat.) Ugyanakkor a szórások meglehetősen magasak, az átlagot meghaladó mértékűek. Ez pedig megnehezíti a pontos becslések készítését, hiszen a standard hibát alapvetően ez a szórás fogja meghatározni.

8. TÁBLÁZAT: SERTÉSÁLLOMÁNY OSZTÁLYKÖZÖS GYAKORISÁGI SORA

Sertésállomány nagyságkategóriái	Gyakoriság (gazdaságok száma)
1-7	45 130
8-15	3 857
16-30	1 538
31-90	646
91-120	43
121-300	59
300-	20
Összesen	51 293

*Forrás: KSH adatbázis alapján, saját szerkesztés*

9. TÁBLÁZAT: A SERTÉSÁLLOMÁNY LEÍRÓ STATISZTIKÁJA KÜLÖNBÖZŐ SOKASÁGOK SZERINT

Mutató	Sertést tartó gazdaságok köre	Állattartó gazdaságok köre	Összes gazdaság
Gazdaságok száma	51 293	79 837	110 949
Átlag	4,41	2,83	2,04
Medián	2	2	0
Módusz	1	0	0
Minimum	1	0	0
Maximum	473	473	473
Terjedelem	472	473	473
Szórás	12,18	9,99	8,57
Ferdeség ( $S_2$ )	0,689	0,645	0,708
Ferdeség ( $S_2'$ )	5,431	4,634	5,849
Értékösszeg	226 054		

*Forrás: KSH adatbázis alapján, saját számítás*

A sertéstartók 92 százaléka legfeljebb 10 állattal rendelkezik, és együttesen a teljes állománynak majdnem a felét adják. Ez azt is jelenti, hogy a sertéstartó gazdaságok maradék



8 százaléknak az állománya adja a hiányzó 50 százalékot. Ez még akár „jól is jöhet” később, mert ezt a viszonylag kis létszámú csoportot koncentrált megfigyelés alá lehet vonni, lefedve ezzel a sertésállomány jelentős részét. A sertésállomány eloszlása közepes mértékű, jobb oldali aszimmetriát mutat. Az  $S_2'$  mutató alapján azt mondhatjuk, hogy a szarvasmarhához, és a következőkben bemutatott juhokhoz képest az aszimmetria mértéke jóval alacsonyabb, illetve kiegyensúlyozottabb képet mutat, ahogy változtatjuk a sokasági szemléletet. Ebből arra is következtethetünk, hogy a sertés az az állatfaj, amelyet a legkevésbé érint az alapsokaságnak tekintett gazdaságcsoport megválasztása. Ennek oka abban keresendő, hogy a sertéstartók száma (több, mint 51 ezer gazdaság) jóval meghaladja a 3-4 ezer közötti juhot, illetve szarvasmarhát tartók számát. Vagyis a sertést nem tartók száma arányaiban sokkal kisebb, mint a másik két állatfaj esetében (ahogy ezt már a 4.3-as fejezetben is bemutattam).

A módusz és a medián értéke is 0-1-2 értékeket vesz fel. Vagyis a tipikus sertéstartó gazdaságban 1-2 állattal találkozunk. Ez a nagy tömegű, viszonylag homogén, néhány állatot tartó gazdaság valószínűleg könnyen megragadható egy kisebb méretű mintán keresztül is. (Lásd: 8. táblázat és 9. táblázat.) A szórásokról ugyanaz mondható el, mint a szarvasmarhák esetében. Megfelelő rétegzés hiányában a szórások ilyen nagyságrendje nagyon megnehezíti a mintából történő becsléseket.

10. TÁBLÁZAT: JUHÁLLOMÁNY OSZTÁLYKÖZÖS GYAKORISÁGI SORA

Juhállomány nagyságkategóriái	Gyakoriság (gazdaságok száma)
1-15	2 653
16-30	384
31-120	287
121-300	202
301-600	106
601-	19
Összesen	3 651

*Forrás: KSH adatbázis alapján, saját szerkesztés*

A juhok esetében is a korábbiakhoz hasonlóan alacsony a tartott állomány módusza (0-2). A ferdeség a szarvasmarhánál látottakhoz hasonlít, de érdekes módon kicsit enyhébb. A sertésekhez képest viszont nagyon magas értékeket látunk. És az erős aszimmetria mellett a tartott állomány nagyság terjedeleme itt a legmagasabb. A legnagyobb, egy gazdaság kezében lévő állomány megközelíti az ezret. Ebből adódóan a sokaság szóródása is itt a legmagasabb, abszolút és relatív értelemben egyaránt. Ráadásul, az előbbi két fajjal ellentétben, itt a teljes

állomány több mint felét a 200 fölötti juhot tartó gazdaságok adják. Vagyis a nagyobb gazdaságok hozzájárulása is jelentős, és még a számosságuk is viszonylag magas. Ez önmagában még nem lenne gond, de az egyes gazdaságok állománya még itt is nagymértékben szóródik. A juhtartás tűnik tehát a leginkább sajátosnak és speciálisnak a három állatfajt összehasonlítva. (Lásd 10. táblázat és 11. táblázat.)

11. TÁBLÁZAT: A JUHÁLLOMÁNY LEÍRÓ STATISZTIKÁJA KÜLÖNBÖZŐ SOKASÁGOK SZERINT

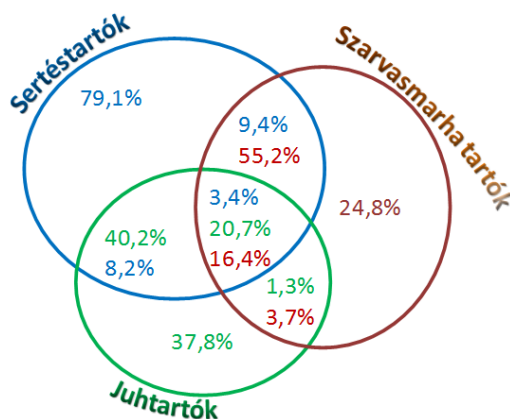
Mutató	Juhot tartó gazdaságok köre	Állattartó gazdaságok köre	Összes gazdaság
Gazdaságok száma	3 848	79 837	110 949
Átlag	38,56	1,86	1,34
Medián	7	0	0
Módusz	2	0	0
Minimum	1	0	0
Maximum	955	955	955
Terjedelem	954	955	955
Szórás	93,73	22,17	18,83
Ferdeség ( $S_2$ )	0,670	0,942	0,950
Ferdeség ( $S_2'$ )	5,061	33,483	39,000
Értékösszeg	148 388		

*Forrás: KSH adatbázis alapján, saját számítás*

A korábbiakban bemutatott halmazok segítségével azt sem érdektelen bemutatni, hogy a gazdaságok száma mellett hogyan alakul a 3 kiemelt fajta tartók állományainak aránya, az átfedéseket is figyelembe véve. A 15. ábraán a különböző színek a különböző állatfajokat jelképezik (kék – sertés, bordó – szarvasmarha, zöld – juh). Minden esetben az adott faj teljes állománya adja a 100 százalékot. Egy-egy állatfaj teljes állománya képezi a viszonyítási alapot (a 100 százalékot), és így látszanak a megoszlások az egyes halmazok, metszetek szerint. Ebből azt olvashatjuk le, hogy a „tisztá” profilú állattartóknál van-e az állomány jelentős része, avagy az jobban eloszlik a különböző csoportok között. Természetesen mintavételi szempontból a kevesebb átlapolás hozna egyszerűbb helyzetet, az empirikus adatok azonban mást mutatnak. A sertéseknél látjuk a az állománynagyság legnagyobb részét a specializálódott gazdaságoknál (79,1%). A legvegyesebb eloszlást a szarvasmarhák produkálják. Az állomány mindössze 1/4-e tartozik a kizárólag szarvasmarhát tartó gazdaságokhoz, és 3/4-e pedig más állatfajjal közös tartásban található. Hangsúlyoznom kell, hogy a dolgozat során a probléma némileg leegyszerűsítve jelentkezik, hiszen csak 3 állatfajt vizsgálunk, figyelmen kívül hagyva a többit. A halmazok száma és az átlapolások, metszetek mennyisége ugrásszerűen növekedne, ha 4-5-6, vagy akár még több állatfajt is bevonnánk (a

földterületi sajátosságokról nem is beszélve). A korábbiakban már bemutatott, több változó egyidejű felhasználása mellett történő rétegzés emiatt a sokszínűség miatt, az átfedések jelentős arányának köszönhetően okoz meglehetősen magas belső szórásokat. Az egyéni gazdaságok jellemzői tehát még ilyen körülhatárolt vizsgálatok esetén is jelentős akadályokat gördítenek a mintavételi tervet kialakító statisztikus elé.

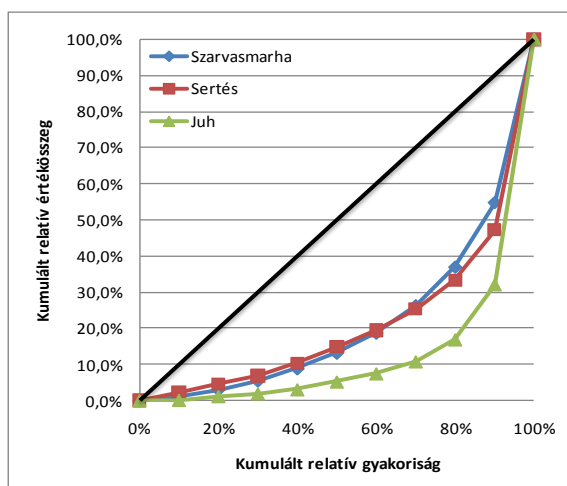
15. ÁBRA: ÁLLATÁLLOMÁNY ARÁNYOK HALMAZOK SZERINT



*Forrás: KSH adatbázis alapján, saját számítás*

A gazdaságok számát és az állomány nagyságának alakulását szemléletesen a Lorenz-görbe (Pintér-Rappai 2007 124. o.) segítségével tudjuk egy grafikonban összefoglalni. A következő (16. ábra) ábrán a három állatfaj koncentrációját együtt ábrázolom, hogy a különbségeket jobban érzékeltetni lehessen. A sokaságokat 10 egyforma részre bontva alakítottam ki az ábrát. Annál jobban koncentrálódik a vizsgált jelenség, minél inkább eltávolodik a görbe a segédvonalként használt átlótól. Azt látjuk, hogy mindhárom állatfaj esetében legalább közepes mértékű koncentrációt találunk. A juhok mutatják a legextrémebb képet (ahogy erre már a leíró adatokból is következtetni tudtunk.)

## 16. ÁBRA: LORENZ-GÖRBÉK



*Forrás: KSH adatbázis alapján, saját számítás*

A háromféle szemléletű sokaságot tekintve mindhárom állatfajnál azt látjuk, hogy a leginkább kezelhető megközelítést az jelentené, ha a legszűkebb értelemben vett, az adott állatfajt tartó gazdaságokra tudnánk fókuszálni a mintavétel és az összeírások során. Az aszimmetria, a szóródás és a középértékek alakulása is ott a legkedvezőbb a mintavételi szempontokat szem előtt tartva. A mintavételi szempontjából viszont egy ilyen lehatárolt szemlélet, és különböző alapsokaságok kezelése akadályokba ütközhet, de erre majd a konkrét szimulációs eredmények bemutatása során fogok kitérni.

### 4.5. Alkalmazott mintavételi módszerek

EU-s részről újabb szabályozások láttak napvilágot. 68 százalékos valószínűségi szint mellett (1 szigma szabály) a hatályos EU-s jogszabálynak<sup>28</sup> megfelelően legalább a következő mintavételi hibahatárokat kell teljesíteni országos szinten a rendszeres éves, illetve évközi összeírások alkalmával az állatállomány szempontjából az értékösszegekre vonatkozóan (12. táblázat):

<sup>28</sup> Az Európai Parlament és a Tanács 1165/2008/EK rendelete [2008. november 19.] az állatállományra és a húsokra vonatkozó statisztikákról, valamint a 93/23/EGK, 93/24/EGK és 93/25/EGK tanácsi irányelvek hatályon kívül helyezéséről

12. TÁBLÁZAT: MAXIMÁLIS RELATÍV STANDARD HIBA ÁLLATFAJONKÉNT AZ EU ELŐÍRÁSOK ALAPJÁN

Állatfaj	Maximális relatív standard hiba (%)
Szarvasmarha	5
Tehén	5
Sertés	2
Juh	2
Kecske	5

*Forrás: Az Európai Parlament és a Tanács 1165/2008/EK rendelete. Saját szerkesztés*

1165/2008/EK 4. cikk, 4. bekezdésében kimondják, hogy a kecskeállományra a pontossági követelmények csak akkor érvényesek, ha legalább 500 ezres állománnyal rendelkezik az ország. Miután Magyarország teljes kecskeállománya 2007-ben nem érte el a 70 ezret sem, ezért a kecske esetében nem kell foglalkoznunk az előírással, és talán nem vállalom nagy kockázatot, ha azt mondom, hogy a jövőben sem fogja elérni az állatszám a kritikus mértéket. A juhok esetében, mivel éppen egy millió felett van az állomány (még 2010-ben is 1 181 ezer az állomány nagysága<sup>29</sup>), a szigorúbb előírásnak kell megfelelni, és 5% helyett 2% az előírt relatív pontosság mértéke.

Mivel csak az egyéni gazdaságok összeírása történik mintavételes módszerrel, ezért a gazdasági szervezetek teljes körűen összeírt adatai tulajdonképpen 0 százalékos mintavételi hibát tartalmaznak. A két gazdálkodói kör számosságának és teljesítményének személtetésére a 13. táblázatban összegyűjtöttem a 2007-es Gazdaságszerkezeti Összeírás (GSZÖ 2007) főbb vonatkozó adatait. Ez azt jelenti, hogy a két gazdálkodó kör arányait felhasználva kellene meghatározni, hogy konkrétan az egyéni gazdaságok esetében mi az az országos mintavételi hibaszint, amit a reprezentatív felvételekkel nem szabadna meghaladni.

<sup>29</sup> Forrás: Statisztikai tükör: Állatállomány, 2010. december 1.

13. TÁBLÁZAT: A KIEMELT ÁLLATFAJOK JELLEMZŐ ADATAI, 2007

Állatfaj	Gazdasági szervezetek száma	Egyéni gazdaságok száma			Állatállomány		
		Összes	Állattartó	Állatfajta tartó	Gazdasági szervezeteknél	Egyéni gazdaságoknál	Országosan
Szarvasmarha	7 656	618 665	338 982	18 907	485 250	219 827	705 077
Sertés				281 930	2 603 958	1 267 189	3 871 147
Juh				21 468	172 660	1 059 345	1 232 005

Forrás: GSZÖ 2007 alapján saját számítás

A kutatás ezen pontján érkeztem el egy komoly dilemmához. Első ránézésre egyszerűnek tűnhet a két gazdálkodói kör adataiból kiszámítani, hogy mik is az egyéni gazdaságok adataitól követelhető hibaszámok. Ugyanakkor mélyebben belegondolva a problémába, a hibaszámok összevonása, összesúlyozása korántsem egyszerű kérdés. Ay (1976) szerint lehetőség van a rétegenkénti standard hibák négyzeteit az értékösszegek négyzetével súlyozni. Az értekezésből ugyanakkor nem derül ki egyértelműen, hogy ez általánosan, vagy csak a hányadosbecslés esetén alkalmazható, illetve hogy milyen megfontolások állnak egy ilyen „nem szokványos” súlyozási módszer hátterében. Mivel a szakirodalomban eddig nem talákoztam hasonló problémafelvetéssel, kénytelen vagyok megelégedni azzal, hogy a *gazdálkodó szervezet – egyéni gazdaság* bontást egyfajta rétegzésnek tekintem, és eszerint járok el a számítások során. Tehát a rétegzett mintavétel szabályai szerint az abszolút hibaszámokat az egyes rétegek elemszámainak, pontosabban azok négyzetének segítségével kell súlyozni, a következő formában:

$$s_y^2 = \sum_{h=1}^L \left( \frac{N_h}{N} \right)^2 \frac{s_h^2}{n_h} \left( 1 - \frac{n_h}{N_h} \right)$$

ahol:

$s_y^2$  - az átlagbecslés standard hiba négyzete

$h$  – a rétegek (1, 2, ..., L)

$N$  – alapsokasági elemszám

$n$  – mintaelemszám

$s$  – rétegenkénti szórás

Esetünkben ugyan értékösszeg becslésről van szó, de ez a formulában érvényesülő összefüggéseket nem érinti, hiszen az átlag és az értékösszeg között csak egy alapsokasági elemszámmal (konstanssal) történő szorzás jelenti az átmenetet.

A formulát átalakítva, és figyelembe véve, hogy csak két rétegünk van, amelyek közül az egyik réteg standard hibája nullával egyenlő, a következő összefüggést kapjuk az egyéni gazdaságok relatív standard hibájára, az előírásokat szem előtt tartva:

$$V = \frac{\sqrt{\frac{N_{eg}^2 s_y^2}{N^2}}}{Y}$$

ahol:

eg – egyéni gazdaságok

V – (előírt) relatív standard hiba

Y – alapsokasági értékösszeg

$s_y$  – értékösszeg becslés standard hibája

Ez alapján természetesen számszerűsíthetjük a teljesítendő hibák mértékét a képlet  $s_y$ -ra való átrendezésével:

$$s_y = \frac{V \cdot Y \cdot N}{N_{eg}}$$

Az így kapott standard hiba értékét kell elosztanunk az egyéni gazdaságoknál található állatállomány létszámmal ahhoz, hogy megkapjuk az egyéni gazdaságokra érvényes relatív standard hibát a különböző állatfajok esetén.

Már ezen a ponton is kérdéses az, hogy milyen alapsokasági elemszámokat használjunk fel. A teljes alapsokaságot, csak az állattartó gazdaságokat vagy esetleg csak az adott állatfaj tartásával foglalkozók számosságát? Nyilvánvalóan az eredményeket ez befolyásolja. A 7. és 8. táblázat tartalmazza a különböző sokasági szemléletek mellett elérendő abszolút és relatív standard hiba értékeit. Ebből azt látjuk, hogy lényeges eltérés csak az állatfajt tartó gazdaságok szemléletében érhető tetten. Ráadásul éppen a legkritikusabb állatfaj (juh)

esetében tapasztaljuk a legnagyobb relatív növekedést, ami akár jó hír is lehet a továbbiakra nézve. Ugyanakkor a háromféle megközelítés végigvonul a különböző mintavételi módszereken is. A szükséges mintanagyságok meghatározásához pedig mindig a konkrét sokasági szemlélethez tartozó hibaszámoknak kell tehát megfelelni. A végső döntéshez azt kell mérlegelni, véleményem szerint, hogy a leghatékonyabb mintavételi terv kialakítása során mit tekintünk alapsokaságnak.

Ha azonban az egyéni gazdaságokat háromféle szemléletben is megközelíthetjük, akkor ugyanezt megtehetjük a gazdasági szervezetekre vonatkozóan is. Újabb kérdésként az merül fel: Milyen szemléleteket kombinálhatunk ebben a speciális esetben, amikor is a gazdasági szervezetek értékösszeg adata egyik szempont szerint sem rendelkezik mintavételi hibával? Elméleti megfontolásból érdekes probléma, de jelen esetben a gazdasági szervezetek száma olyan mértékben kicsi az egyéni gazdaságokéhoz képest, hogy a számítások végeredménye szempontjából nincsen jelentősége a különbségtételnek, ezért el is tekintek ettől.

14. TÁBLÁZAT: MEGENGEDHETŐ MAXIMÁLIS ÉRTÉKÖSSZEGRE VONATKOZÓ HIBANAGYSÁGOK AZ EGYÉNI GAZDASÁGOKNÁL

Állatfaj	A maximális standard hiba értéke a következő sokaságot véve alapul:		
	Összes gazdaság	Állattartó gazdaságok	Állatfajt tartók
Szarvasmarha	35 671	36 024	49 061
Sertés	78 339	79 114	85 337
Juh	24 932	25 178	33 139

*Forrás: KSH adatbázis alapján saját számítás*

15. TÁBLÁZAT: MEGENGEDHETŐ MAXIMÁLIS RELATÍV HIBANAGYSÁGOK AZ EGYÉNI GAZDASÁGOKRA VONATKOZÓAN (%)

Állatfaj	Az egyéni gazdaságok állományára vonatkozó maximális relatív hibanagyságok a következő sokaságot véve alapul:		
	Összes gazdaság	Állattartó gazdaságok	Állatfajt tartók
Szarvasmarha	16,04	16,39	22,32
Sertés	6,18	6,24	6,73
Juh	2,35	2,38	3,13

*Forrás: KSH adatbázis alapján saját számítás*



Kish (1995b) további súlyozási lehetőségeket mutat be, azok azonban minden esetben valamilyen hányados formájában kiszámítható mutatók becslésére vonatkoznak. Tehát ezt figyelmen kívül is hagyhatom a jelenlegi problémát tekintve.

A korábbi évek hibaszámai (16. táblázat) azt mutatják, hogy a szarvasmarha, és a sertés esetében még ez az „öszvér” minta is képes (többé-kevésbé) biztosítani az előírt hibaszintet. Ebből arra következtetek, hogy egy ésszerűbb, specializált minta jóval alacsonyabb mintaelemszám mellett tudna megfelelő hatékonyságot elérni. Sajnos a juh esetében viszont túlhaladják a hibaszámok az elvárt nagyságrendet. Ez pedig azt jelenti, hogy a gyakorlatban jelenleg alkalmazott univerzális mintavételi terv nem fog kielégítő eredményre vezetni lényeges módosítások bevezetése nélkül.

16. TÁBLÁZAT: EGYES KORÁBBI IDŐSZAKOK HIBASZÁMAI

Állatfaj	Az országos állomány relatív standard hibája (%)					
	2004. XII. 1.	2005. VIII. 1.	2006. XII. 1.	2007. VIII. 1.	2008. XII. 1.	2009. XII. 1.
Szarvasmarha	2,44	2,59	2,28	1,96	1,93	2,11
Sertés	2,28	2,13	1,3	1,22	1,66	1,49
Juh	11,33	11,15	9,98	9,7	11,65	12,09

*Forrás: Állatállomány, 2004. április 1., Állatállomány 2005. augusztus 1., Állatállomány 2006. december 1. alapján saját számítás*

A 2010-es cenzust követő első reprezentatív összeírás a decemberi állatállomány megfigyelés és az éves összeírás volt. Ehhez már ki kellett volna alakítani az új mintavételi módszertant. Azonban a nyári összeírás, és az őszi mintaválasztás közötti idő kevésnek bizonyult erre. Ezért azt az áthidaló megoldást választották, hogy az ÁMÖ úgynevezett mesterkörzeteit használták fel. A census címlistája alapján keresték fel az egyéni gazdaságokat ezekben a körzetekben. Ezzel egy körülbelül 70 ezres mintasokaság megfigyelésére került sor.

A mintavételi hiba a pontosságnak csak az egyik részét adja. Fontos látni, hogy felmerülhetnek az összeírás során nem mintavételi hibák is. Ennek részletes bemutatását a 3.3 fejezet tartalmazza. Ugyanakkor, mivel az EU-s jogszabályok a pontosságot a standard hiba nagyságán keresztül határozzák meg, a későbbiekben a dolgozat is csak ezzel foglalkozik.

Miután vizsgálataim alapjaként tehát a mintavételi hibát tekintem a legszigorúbb feltételnek, az alapsokasági szórások vizsgálatával kell kezdeni a számításokat. Ugyan a becslések értékösszegre vonatkoznak, az egyszerűség kedvéért a vizsgálatokat az átlagokra alapozva végzem el. A relatív szórás és a szükséges mintanagyság szempontjából ugyanazt az eredményt kapom így is, hiszen az értékösszeg egyenlő az átlag és az alapsokasági elemszám szorzatával.

A szórás és átlagszámítás „érdekességét” az adja, hogy a KSH adatbázisában az állatállományi adatokat tartalmazó táblában csak azok a gazdaságok szerepelnek, amelyek ténylegesen tartanak valamilyen állatot az összeírás referencia-időpontjában. Vagyis az alapsokaságra jellemző mutatókat akkor kapjuk meg, ha a számítások alapját kiegészítjük a csak növénytermesztéssel foglalkozó, illetve az adott állatfaj tartásával nem foglalkozó, de egyébként állattartó gazdaságok üres, vagyis nullás adataival is. Ez az átlagot nyilvánvalóan csökkenteni fogja, a szórást pedig növelni. Már itt érdekes, bár nem hoz meglepő eredményt a különböző szemléletű sokaságok kialakításával kapható szórásértékek különbségének a vizsgálata.

17. TÁBLÁZAT: A RELATÍV SZÓRÁS ALAKULÁSA SOKASÁGOK ÉS FAJOK SZERINT (%)

Állatfaj	Összes gazdaság	Állattartó gazdaságok	Állatfajt tartó gazdaságok
Szarvasmarha	981,6	831,0	142,3
Sertés	420,5	352,8	276,4
Juh	1407,6	1192,9	243,0

*Forrás: KSH adatbázis alapján saját számítás*

A 17. táblázat tartalmazza a három különböző megközelítésű sokaságra vonatkozóan a relatív szórás értékeket. Azt szokták mondani, hogy a relatív szórás 100 százalék felett már szinte értelmét veszti. Nos, az állatállományok szempontjából sajnos nagyon extrém eredményeket találunk. Különösen az első két változat esetében tapasztalunk rendkívül magas értékeket. Ez jól jellemzi az állatfajok heterogenitását. Amit a számok ismerete nélkül is feltételezni lehetett – nevezetesen, hogy a szűkített sokasági értelmezés homogenizáló hatású lesz – azt az eredmények is igazolják. Talán annyi volt meglepő a számomra, hogy a relatív szórások a teljes sokaság és az állattartók között viszonylag keveset csökkentek. Természetesen a legideálisabb verziót az állatfajt tartók eredményei adják, ugyanakkor ez a fajta megközelítés okozhatja a legtöbb problémát az átfedések miatt.

A továbbiakban, a szórások ismeretében sorra veszem a korábban bemutatott mintavételi módszereken. Először külön-külön megvizsgálva, hogy az egyes módszerek önálló alkalmazásával mekkora mintaelemszám szükséges az előírt hibahatárok betartásához, majd a különböző módszerek kombinálására is sor kerül, az eredmények alapján. Visszatevés nélküli módszereket alkalmazok, hiszen a gyakorlatban nem lenne sok értelme a visszatevéses – tulajdonképpen inkább elméleti szempontból fontos – eljárást választani.

## 5. Szimulációs eredmények

Ebben a fejezetben részletesen bemutatásra kerülnek a kísérleti adatbázis alapján készített szimulációs számítások és azok eredményei. Sorra veszem a különböző mintavételi módszerekkel elérhető eredményeket. A csoportos kiválasztást nem modellezem, hiszen korábban (a 3.1.3 fejezetben) már bemutattam, hogy a mezőgazdaság statisztikai összeírásoknál a sokaság jellege miatt nem számít hatékony módszernek ez a technika. Ezért a KSH-ban jelenleg alkalmazott mintavételi tervet és annak eredményeit csak a végső összehasonlításhoz, a számvetésnél fogom felhasználni.

Nagyon fontosnak tartom tisztázni a fejezet elején, hogy mit értek szimuláció alatt. Mivel a GSZÖ 2007 mintasokaságot alapsokaságnak tekintem, ezért az így számított szórásokat fogom tudni felhasználni a szükséges mintaelemszámok kiszámításához. A gyakorlatban sokszor okoz problémát az a tény, hogy az alapsokasági szórásokat előre nem ismerjük, és általában még közelítő információval sem rendelkezünk a nagyságukról az összeírás lebonyolítása előtt. Ezt a problémát hidalom át azzal, hogy a szórásokat ismertnek tételezem fel, felhasználva ezt a viszonylag nagy arányú és elemszámú mintasokaságot.

A számokkal és a belőlük levonható következtetésekkel párhuzamosan statisztika-módszertani megfontolások kiemelésére, helyenként újszerű megközelítések ismertetésére is sor kerül ebben a fejezetben.

### 5.1. Egyszerű véletlen mintavétel

A legegyszerűbb módszer az egyszerű véletlen mintavétel. A szórások ismeretében megadható, hogy mekkora az a mintanagyság, ami az előírt hibanagyságokat teljesíti. Az egyszerű véletlen (valószínűségi) mintára az a jellemző, hogy valamennyi alapsokasági elem mintába kerülési valószínűségét ismerjük már a mintavétel elején, és ez a valószínűség minden elem esetében egyenlő és nagyobb, mint nulla. Ekkor a következő formulával kapjuk meg a szükséges mintanagyságot (tekintve, hogy 68 százalékos megbízhatósági szint melletti deltát használunk):

$$n = \frac{\sigma^2}{\Delta^2 + \frac{\sigma^2}{N}}$$

ahol:

n: mintanagyság

N: alapsokaságnak tekintett csoport teljes elemszáma

$\Delta$ : elvárt hibahatár (konfidencia intervallum)  $(V_{eg} \times \bar{y})^{30}$

A különböző szempontok szerint számított szórásoknak megfelelően szükséges kiválasztási arányokat visszatevés nélküli, egyszerű véletlen kiválasztást feltételezve, a következő táblázat foglalja össze.

18. TÁBLÁZAT: ÁLLATFAJONKÉNT SZÜKSÉGES EV MINTAVÉTELI ARÁNY, KÜLÖNBÖZŐ SZEMLÉLETŰ ALAPSOKASÁG MELLETT (%)

Állatfaj	Teljes alapsokaság	Állattartó gazdaságok	Az állatfajt tartó gazdaságok
Szarvasmarha	3,27	3,1	1,2
Sertés	4,09	3,9	3,6
Juh	76,69	75,9	61,0

*Forrás: KSH adatbázis alapján saját számítás*

A 18. táblázat az adott alapsokaság arányában tartalmazza a szükséges mintavételi arányt. A relatív szórások csökkenéséből is látható volt, hogy az alapsokaság értelmezésének szűkítése nyilvánvalóan javítani fogja a hatékonyságot, vagyis kisebb mintavételi aránnyal is elérhető az előírt hibahatár. Ahogy a szórások esetében is láttuk, a teljes és az állattartó gazdák sokasága között ilyen szemléletben nincsen jelentős eltérés, viszont a konkrét állatfajt tartók köre ezekhez képest lényegesen jobb százalékokat mutat.

<sup>30</sup> Mivel 68 százalékos megbízhatósági szint mellett vizsgáljuk a szükséges mintanagyságot, ahol pedig a konfidencia intervallum becslésénél a normális eloszlásból származó érték 1, így ez nem módosítja a delta értékét, ezért nem is szerepeltetem a képletben.

A szarvasmarha és a sertés esetén már kiinduláskor is elfogadható mintavételi arányokat kaptunk eredményül. Viszont a juh esetében ránézésre azt mondhatjuk, hogy a 3. megoldás lehet egy elfogadható mintavételi terv alapja, hiszen annyira speciális állattartókról van szó, hogy mindkét másik esetben az alapsokaság  $\frac{3}{4}$ -edét meg kellene vizsgálni a kívánt pontosság eléréséhez. Hunyadi (2001b) mutat be egy hasonló példát, ahol 3-féle szántóföldi terület búzatermésének átlagához keresi az optimális mintanagyságot. A példában is az a helyzet áll fenn, hogy az egyik terület viszonylag kis súllyal rendelkezik, ugyanakkor a termésátlagok szórása itt a legnagyobb. Ebben az esetben, a súlyok és a szórások figyelembe vétele mellett ez a terület kiugróan nagy szükséges mintaelemszámot „kívánna”. Ugyanakkor, mivel a súlya a teljes területen belül kicsi, Hunyadi javaslata az, hogy ilyen esetekben meg kell elégedni nagyobb szórással is a kisebb súlyú területre vonatkozóan, hogy ne okozzon feleslegesen nagy mintát. A mi állatösszeírásainkra lefordítva ez azt jelentené, hogy mivel a magyar állattartásban a juhtartás nem képvisel jelentős súlyt, ugyanakkor a gazdaságoként előforduló állománynagyság rendkívül változatos (nagy szórással rendelkezik), elégedjünk meg nagyobb hibaszámokkal is az állatfajnak a becslése során, hogy a mintanagyság ezek miatt ne nőjön túlságosan (indokolatlanul) nagyra. Ha belegondolunk, ez nem is egy földtől elrugaszkodott elképzelés, hiszen, ami nem olyan jelentős az országban, azt ne is akarjuk olyan nagy biztonsággal megbecsülni. Igen ám, de az EU-s előírások világosan meghatározzák, hogy mi az a hibaszám, amivel megelégedhetünk, így ez a fajta megoldás nem játszhat szerepet, mást kell kitalálni. Mindemellett elgondolkodtató az EU-s előírások létjogosultsága, illetve a hibahatár teljesítéséhez meghatározott országos állománynagyságok kijelölésének szakmai megalapozottsága. Lehetséges, hogy az egyes állatfajokhoz nem is nagyság, hanem az adott ország mezőgazdaságában betöltött jelentőség, illetve valamilyen érték (pl.: SFH) alapon kellene módosítani az előírásokat. Ezzel jobban lehetne alkalmazkodni az egyes országok sajátosságaihoz, és összességében, a fontosabb állatfajokra vonatkozó adatok minősége sem sérülne.

Ahhoz, hogy a szükséges minták nagyságrendjét is össze tudjuk hasonlítani, a 19. táblázatban a teljes alapsokasághoz, vagyis az összes gazdasághoz viszonyítva látjuk a szükséges mintavételi arányokat. Természetesen az első oszlop értékeiben ez nem hoz változást, hiszen ott eredetileg is az összes gazdaság volt a viszonyítási alap. Azt látjuk a számokból, hogy az első két sokaság a juh esetére még mindig túlságosan magas arányokat jelent, viszont az utolsó oszlop már figyelemre méltó. A konkrét állatfajjal rendelkező gazdák

számossága eleve kisebb, és az összes gazdasághoz viszonyítva nagyon barátságosnak tűnnek ezek az arányok.

19. TÁBLÁZAT: SZÜKSÉGES MINTAVÁLASZTÁSI ARÁNYOK A TELJES ALAPSOKASÁGHOZ  
VISZONYÍTVA (%)

Állatfaj	Teljes alapsokaság	Állattartó gazdaságok	Az állatfajt tartó gazdaságok
Szarvasmarha	3,27	2,24	0,04
Sertés	4,09	2,77	1,69
Juh	76,69	54,60	3,45

*Forrás: KSH adatbázis alapján saját számítás*

A szarvasmarha és a sertés esetén felmerülhet bennünk az a kérdés, hogy miért nem állunk meg rögtön ezen a ponton. Egy 2-4 százalékos mintavételi arány a gazdaságstatisztikai összeírásokban teljes mértékben elfogadható nagyságrendet képvisel. A számossággal nincs is probléma. Viszont az egyszerű véletlen mintavétel sajátossága, hogy területi, összeírásszervezési, költségvetési szempontokat egyáltalán nem érvényesít. A mintasokaság elemei területileg nagyon szétszórtnak jelentkeznének, ami bonyolultabbá, drágábbá és időigényesebbé tenné a rendszeres összeírásokat. Ezért tehát ennél alacsonyabb mintavételi arányra lesz szükségünk, ha az összeírókat szétszórtnak címekre szeretnénk kiküldeni. A másik ok, amiért az egyszerű véletlen nem lehet optimális megoldás, az a területi reprezentativitás igénye. Az EV mintavételi terv részleteit elsősorban azért kellett a vizsgálatok elején bemutatni, mert ehhez minden további módszert viszonyítani tudunk majd. Ez mintegy támpontot ad a többi mintavételi terv értékelésekor.

## 5.2. Rétegzett mintavétel

Elérkeztünk a rétegzett mintavételhez. A rétegzés alkalmazásának alapfeltétele, hogy teljes körű információval rendelkezünk a rétegző ismérvről az alapsokaságra vonatkozóan. Fontos eleme a rétegzett mintavételnek, hogy minden sokasági elem csak egy rétegbe kerüljön bele, és minden elem besorolható legyen. A korábban már tárgyaltaknak megfelelően a sokasági egységek természetéből fakadóan, illetve a rendelkezésre álló regiszterinformációk miatt megkülönböztetjük az egyéni gazdaságok és a gazdasági szervezetek csoportját. Ez természetesen nem jelent esetemben rétegzést, hiszen csak az egyéni gazdaságokat vizsgálom. Ugyanakkor a teljességhez ez az aspektus is hozzátartozik.

Rétegzés tervezésekor számos kérdést kell feltennünk és megválaszolnunk: Mi a legjobb rendelkezésre álló rétegeképző ismérv? Hogyan húzzuk meg a rétegek közötti határokat? Hány réteget alakítsunk ki? Ezeket a kérdéseket igyekszem a későbbiekben magam is megválaszolni.

Optimális esetben az lenne a legjobb rétegeképző ismérv, amire a megfigyelés irányul. Erről azonban sosem áll rendelkezésünkre teljes körű, pontos információ, hiszen éppen ezt szeretnénk begyűjteni. (A későbbiekben bemutatásra kerül az úgynevezett kétfázisú mintavétel módszere, amely minden körülmények között lehetővé teszi a vizsgált ismérv szerinti rétegzést.) Az állatállomány összeírások szempontjából különösen nehéz olyan rétegeképző ismérve(ke)t találni, amelyekre igaz az, hogy minden elem csak egy csoportba tartozik bele (a több állatfajt tartó gazdaságokat nem lehet csak egy állatfaj szerint kialakított rétegbe illeszteni). Az is nehézséget okoz, hogy nem egy változó szerint szeretnénk optimális eredményt kapni, hanem több változó képezi a vizsgálat tárgyát. Ezért alakult ki az a gyakorlatban, hogy a rétegeképző ismérv meghatározása során többféle feltételt fogalmaznak meg egyszerre, amelyek segítségével a „kisebb” és „nagyobb” gazdaságokat próbálják meg elkülöníteni egymástól. Ezáltal átfedések nem keletkeztek az egyes rétegek között, viszont a lényeg látszik elveszni, vagy legalábbis sérülni ezáltal. Hiszen, ha egy gazdaság egy szempont alapján „nagy” minősül, attól még a többi vizsgált mutatója lehet egészen alacsony (a gyakorlatban jellemző az, hogy a gazdaságok bizonyos mértékig specializálódnak, ezért nagyon is jelentős azoknak a száma, akik egy mutató alapján kimagaslóak vagy legalábbis viszonylag nagyok, a többi alapján azonban a kisebb kategóriába tartoznának). A rétegzésnek pedig az én szemléletem szerint itt elsősorban az a célja, hogy homogén csoportokat alakítsunk ki a vizsgált ismérvek szempontjából, vagyis a rétegeken belüli elemek egymáshoz minél inkább hasonlítsanak és kapcsolati rétegzést hajtsunk végre. Ez a tulajdonság pedig jótékony hatással lenne a mintavételi hiba alakulására, hiszen a rétegeken belüli szórások az alapsokasági szóráshoz képest jóval kisebbek lennének. A fenti, többmutatós rétegeképzésnek azonban az a hátránya, hogy a szórások nagyságát nem képes a kellő mértékben csökkenteni. Ez a problémák egyik forrása.



20. TÁBLÁZAT: A JELENLEG ALKALMAZOTT RÉTEGZÉS SZERINT AZ EGYÉNI GAZDASÁGOK MEGOSZLÁS A 2007-ES ADATOK ALAPJÁN (%)

Rétegek	Állatállomány szerint „A”	Állatállomány szerint „B”
Földhasználat szerint „A”	2,7	6,2
Földhasználat szerint „B”	4,1	86,9

*Forrás: KSH adatbázis alapján saját számítás*

A kétféle szempont (földhasználat vagy állattartás) szerint egy-egy gazdaság tartozhat különböző nagyságkategóriákba is. (Az „A” és „B” réteg definícióját lásd: 3.2.1 fejezet.) Az arányokat a 20. táblázat szemlélteti. Ez nagyvonalakban azt mutatja be, hogy a gazdaságok nem egészen 3 százaléka az, aki igazán nagyinak mondható mindkét szempont szerint. A nagy tömeg pedig minden szempontból kicsi. A maradék 10 százalék az, aki nagy kategóriába kerül, holott csak az egyik termelési típus szerint éri el a küszöböt. Ez egyrészt azt jelenti, hogy legalábbis ez a 10 százaléknyi gazdaság biztosan heterogénebbé teszi az „A” réteget. Másrészt pedig azt, hogy a rétegzés után is a gazdaságok egy nagy tömege „együtt” maradt a „B” rétegben. Hogyha nem csak a gazdaságszám, hanem a mögöttük álló érték alapján is megvizsgáljuk ugyanezt a felbontást (21. táblázat), akkor a „B” réteg kevésbé tűnik olvasztótégelynek, ugyanakkor a félig nagy, félig kicsi gazdaságok pedig értékben még nagyobb súlyt képviselnek.

21. TÁBLÁZAT: A JELENLEG ALKALMAZOTT RÉTEGZÉS SZERINT AZ EGYÉNI GAZDASÁGOK SFH ÉRTÉKÉNEK MEGOSZLÁSA A 2007-ES ADATOK ALAPJÁN (%)

Rétegek	Állatállomány szerint „A”	Állatállomány szerint „B”
Földhasználat szerint „A”	6,7	9,6
Földhasználat szerint „B”	30,9	52,9

*Forrás: KSH adatbázis alapján saját számítás*

A KSH-ban alkalmazott rétegzésnek az is egy jelentős problémája, hogy a census idején fennálló állapotnak megfelelően alakítja ki a rétegekbe sorolást. Ennek könnyen belátható oka az, hogy a gazdaságok jelentős részéről nincsen ennél frissebb információ, így meg kell elégednünk az egyre elavultabbá váló csoportosítással. A 22. táblázat ennek a torzító hatását kívánja számokban kifejezve érzékeltetni. 2000 és 2007 között a csak a 2007-es GSZÖ mintájában szereplő gazdaságok esetében nagyon jelentős eltérés mutatkozik a rétegzés alakulásában. Ezek a nagyságrendi különbségek már tulajdonképpen azt sugallják, hogy az

ÁMÖ-től távolodva az 5-10 éves adatok alapján történő rétegzés hatékonyság javító ereje nagyon lecsökken, legalábbis ami a kapcsolati rétegzést illeti. A rétegekbe sorolás egyre inkább pontatlanná, megbízhatatlanná válik.

22. TÁBLÁZAT: A RÉTEGZÉS VÁLTOZÁSA IDŐBEN

	ÁMÖ 2000 alapján	GSZÖ 2007 alapján
Réteg	Gazdaságok száma	
"A"	29 805	17 547
"B"	60 073	101 038
Nem besorolt	22 152	-

*Forrás: KSH adatbázis alapján saját számítás*

A másik probléma az, hogy az állatállomány összeírásoknak a hazai felhasználók igényei, és az Európai Unió előírások alapján legalább régiós szintre is reprezentatívnak kell lenniük. Ez azt jelenti, hogy kénytelen-kelletlen egy rétegzési szempont eleve adott. A régiók szerint kialakult 7 rétegre külön-külön kell tekinteni, technikai rétegzést kell alkalmazni. További szempontokat csak emellé lehet összegyűjteni, azonban szem előtt kell tartani azt is, hogy a rétegek száma bizonyos „ésszerű” határok között kell, hogy mozogjon. Mivel a rétegzési szempontok által képzett csoportok száma szorozódik, viszonylag kisszámú további réteg kialakítására marad lehetőség.

Itt tartom fontosnak hangsúlyozni, hogy a „régiós szintre reprezentatív” mintavételi terv önmagában nem biztosítja azt, hogy régiós szinten is alacsony hibahatár mellett kapjunk becsléseket a vizsgált paraméterekre. Ez mindössze annyit jelent, hogy az alapsokaság régiós eloszlását a mintavétel során nem lehet mellőzni és fel kell használni. Amennyiben régiós szintű adatokra is elvárt bizonyos alacsony standard hiba, az egészen más megközelítést igényel. (A témában lásd: Galambosné Tiszberger 2010.) Ettől függetlenül természetesen számíthatóak a régiós szintű adatokra vonatkozó standard hibák is, amelyek végül segítenek a felhasználónak eldönteni, hogy számára mennyire megbízhatóak az adatok, és azokat mire tudja felhasználni.

Az alapsokaság kétféle szemlélete végighúzódik a dolgozat egészén. Ezt azért tartom fontosnak, mert indokolatlannak tűnik „egy kalap alá venni” az eltérő profillal (földterülettel, illetve állatállománnyal foglalkozó) rendelkező gazdaságokat, és ezzel mintegy felhígítani a sokaság összetételét, heterogénebb összképet hozva létre.

A szűkítéssel ugyanakkor egyértelműen kettéválasztjuk a földterülettel kapcsolatos összeírások alapját az állatállománytól. Az univerzális minta ezzel megszűnne. Ez hátrányt jelent a lebonyolítás megszervezésében és a költségek alakulásában is. Éppen ezért csak jelentős hatékonyság javulás mellett lehet csak ez a jó megoldás. Még a szimulációs számítások elvégzése előtt felételeztem azt, hogy **amennyiben szűkítjük (pontosítjuk) a teljes, mezőgazdasággal foglalkozó egyéni gazdaságok alapsokaságát** aszerint, hogy rendelkeznek-e egyáltalán állatállománnyal, **az lényegesen javítja** majd **a** mintavételből nyerhető **becslések hatásosságát**. (H1) Megfordítva ez annyit tesz, hogy ugyanolyan mintavételi technika mellett a paraméterbecslés megcélzott relatív standard hibáját lényegesen alacsonyabb mintaelemszámmal, illetve mintavételi aránnyal lehet biztosítani.

Többek között ennek a H1 hipotézisnek a vizsgálatára szolgálnak majd a számítások eredményei.

Az állatfajok szerinti specializációnak megfelelő alapsokasági szemlélet bemutatásának inkább csak azért volt jelentősége, hogy lássuk, ez mennyire homogenizálja az egyedeinket. A továbbiakban azonban nem vizsgálom ezt, mint tényleges gyakorlati lehetőséget. Ennek oka az, hogy a gazdaságok pontos csoportosítása ilyen módon szinte teljesen megoldhatatlan. Ráadásul nem csak a három, jogszabályok által kiemelt állatfajra keresünk értékösszeg becsléseket, hanem a többi haszonállatot (baromfi, kecske, ló, stb.) is meg kell figyelni az összeírások alkalmával. Ha tehát legszűkebb értelemben is az állattartó gazdaságokból indulunk ki, azzal még nem mondunk le a többi állatfajról, és meghagyjuk azok megfigyelésének a lehetőségét is.

### 5.2.1. Rétegzés régiók szerint

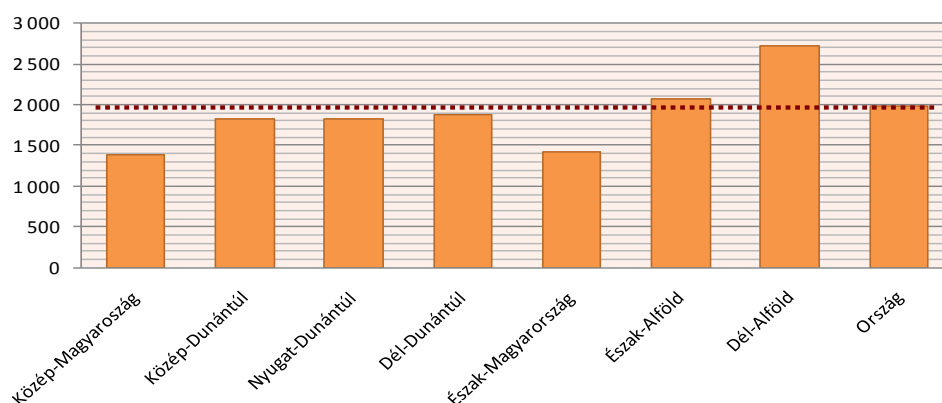
A korábbiakban bemutatott szempontok miatt régiós szintű adatokra feltétlenül szükség van a mezőgazdaság statisztika adatainak publikálásakor, ezért a régió, mint rétegtképző ismérv adottság. Előzetesen azt a hipotézist állítottam fel ezzel kapcsolatban (H2), hogy a **régiós rétegzés nem javít a mintavételi terv hatásosságán** az egyszerű véletlenhez képest, és csak technikai rétegzés valósul meg. Ennek a hipotézisnek a bizonyításához egyrészt meg kell néznünk a régió, mint rétegtképző ismérv kapcsolatát a vizsgált változókkal, másrészt pedig számszerűen is ki kell mutatni, hogy ugyanazt a mintavételi hibát nem tudjuk jelentősen alacsonyabb mintaelemszámmal elérni akkor, ha területi alapon rétegzünk.

A régió, mint rétegeképző ismérv, akkor lenne szerencsés választás (egy pillanatra eltekintve attól, hogy régiós szintű adatokat is szeretnénk becsülni), ha legalább közepes, vagy még inkább erős sztochasztikus kapcsolatot találnánk a gazdaságok régiók szerinti hovatartozása, és az állatállomány, gazdaságméret, gazdaságtípus változók között. Első lépésben ilyen típusú kapcsolatok létezését ellenőriztem. Mivel ebben az esetben minőségi és mennyiségi ismérvek viszonyáról van szó, ezért a vegyes kapcsolat mérésére alkalmas szóráshányados<sup>31</sup> mutató alkalmazható a kapcsolatszorosság méréséhez. A gazdaságméretet az SFH értékkel szemléltetem, mivel ez minden gazdaságra rendelkezésre áll, és egy mutatóba sűrítve jeleníti meg a gazdaság teljesítményét (méretét). A három gazdaságszerkezeti összeírás (2003, 2005, 2007) szerinti adatokat használtam fel. A varianciaanalízishez tartozó F-próba szignifikancia értéke minden esetben gyakorlatilag 0, vagyis van sztochasztikus kapcsolat a két vizsgált ismérv, az SFH és a régió között (nem függetlenek egymástól). Ugyanakkor a kapcsolat szorosságát mérő szóráshányados értéke 0,05 az egyéni gazdaságokra nézve, a gazdasági szervezetek esetében 0,09 mindhárom időszakban. Ez pedig azt jelenti, hogy a sztochasztikus kapcsolat a gazdaság földrajzi elhelyezkedése és mérete között nagyon gyengének mondható. Ez a kapcsolati rétegzéshez biztosan kevés. Elméletileg persze, a szignifikáns kapcsolat léte már hoz hatékonyság javulást, de mértékét tekintve, ahogy ezt a későbbiekben a konkrét számításokban látni fogjuk, ez elenyésző a gyenge intenzitás miatt.

---

<sup>31</sup> Szóráshányados:  $H = \frac{\sigma_K}{\sigma}$ , vagyis a külső és a teljes szórás hányadosa.

17. ÁBRA: AZ EGYÉNI GAZDASÁGOK ÁTLAGOS SFH ÉRTÉKEINEK ALAKULÁSA RÉGIÓNKÉNT, 2007 (€)<sup>32</sup>



*Forrás: KSH adatbázis alapján saját számítás*

Az átlagos (egy gazdaságra jutó) SFH értékek régiónkénti alakulását a 17. ábra és a 18. ábraán láthatjuk. Nyilván nem meglepő az eredmény. A számítások előtt sem gondoltuk, hogy egyes régiókban nagyobb, másikatban pedig jelentősen kisebb gazdaságok működnek, különösen, ha azt is számításba vesszük, hogy a régiós besorolás a gazdálkodó címe, illetve a szervezet székhelye alapján kerül kialakításra. Tehát, ha adottságaikban, a bennük rejlő lehetőségekben mutatnának is lényegesen eltérő képet az egyes régiók, még az sem derülne ki feltétlenül a székhely szerinti besorolás miatt.

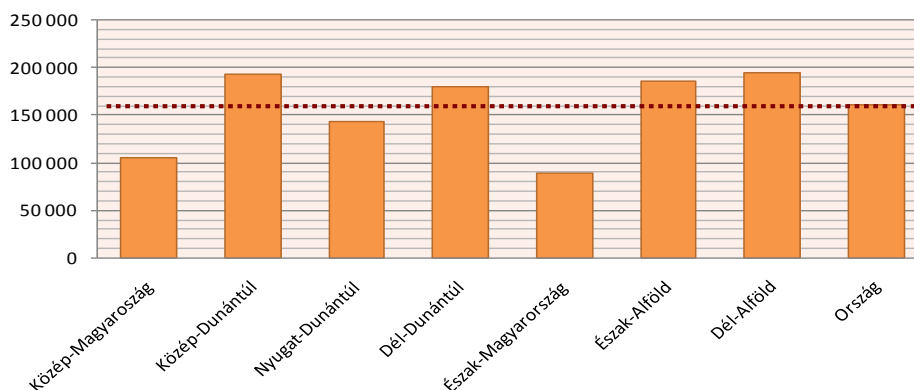
A gazdasági szervezetekre vonatkozó 18. ábra azt mutatja, hogy ebben a körben nagyobb különbségek léteznek a régiók között. A konkrét számítások elvégzése megmutatja, hogy itt sincs még közepes erősségű sztochasztikus kapcsolat sem az elhelyezkedés és a gazdaság méret között.

A régiós csoportosítást tovább vizsgáltam aszerint, hogy a gazdaság méretén kívül a gazdaságok típusával mutat-e valamilyen kapcsolatot. (A gazdaságok régiók és gazdaságtípológia szerinti csoportosítását lásd V. számú melléklet.) Ebben az esetben minőségi ismérvek közötti kapcsolatot vizsgálunk (asszociációs kapcsolat), ezért a Khi-

<sup>32</sup> A szaggatott egyenes vonal az egyéni gazdaságokra jellemző országos átlagot mutatja.

négyzet próbát<sup>33</sup> kell alkalmazni. Ez szintén szignifikáns kapcsolatot mutat a két ismerv között, ugyanakkor, ha megnézzük a Cramer-féle asszociációs együttható<sup>34</sup> értékét (0,1), akkor azt is láthatjuk, hogy a kapcsolat ebben az esetben is meglehetősen gyenge a két változó között.

18. ÁBRA: A GAZDASÁGI SZERVEZETEK ÁTLAGOS SFH ÉRTÉKEINEK ALAKULÁSA RÉGIÓNKÉNT, 2007 (€)<sup>35</sup>



Forrás: KSH adatbázis alapján saját számítás

Mivel a dolgozat elsősorban az állatállománnyal foglalkozik, a három kiemelt faj szempontjából is ellenőriztem a sztochasztikus kapcsolatok létét és erősségét a varianciaanalízis segítségével. A szóráshányados mutató a következők szerint alakul: szarvasmarha 0,03; sertés 0,09; juh 0,04. (A kapcsolat minden esetben szignifikáns.) Tehát a kapcsolat intenzitása a konkrét változókra is meglehetősen gyenge. (Az egy gazdaságra jutó átlagos állomány nagyságokat régióként a VI. számú melléklet tartalmazza.)

Ezek alapján tehát a bevezetőben megfogalmazott **2. hipotézis igazolódni látszik**, vagyis **szoros sztochasztikus kapcsolat hiányában a régiós rétegzés nem fog tudni jelentősen**

<sup>33</sup> Khi-négyzet:  $H_0 : \chi^2 = 0$ ;  $H_1 : \chi^2 > 0$ ;  $\chi^2 = \sum_{a=1}^s \sum_{b=1}^o \frac{(f_{ab} - f_{ab}^*)^2}{f_{ab}^*}$  (Pintér-Rappai, 2007 378. o.)

<sup>34</sup> Cramer-féle asszociációs együttható:  $C = \sqrt{\frac{\chi^2}{n \times \min[s-1, o-1]}}$  (Pintér-Rappai 2007 233. o.)

<sup>35</sup> A szaggatott egyenes vonal a gazdasági szervezetekre jellemző országos átlagot mutatja.

**javítani a mintavételi terv hatásosságán.** A pontos eredményekhez azonban további számításokat kell elvégezni, összehasonlítva az EV és a régiósan rétegzett minta eredményeit.

A régiós rétegzésnél a mintaelemszámok meghatározásához a régió szintjére is szükséges számszerűsíteni a szórást minden vizsgált állatfajra. Ezután nyílik lehetőség arra, hogy megvizsgáljuk, a megcélzott hibahatároknak megfelelően mi az a mintavételi arány, ami megfelelő. A konkrét számításokhoz először azt is meg kell határozni, hogy az egyes rétegek között hogyan oszlik el a teljes mintanagyság. Először az arányos rétegzést vizsgáltam. Ekkor a szükséges mintaelemszámot a következő összefüggés<sup>36</sup> alapján határozhatjuk meg:

$$n = \frac{N \sum_{h=1}^L N_h \sigma_h^2}{(N \frac{\Delta}{u})^2 + \sum_{h=1}^L N_h \sigma_h^2}$$

Mivel „u” értéke esetünkben 1 (68 százalékos megbízhatósági szint mellett), így a képlet valamelyest egyszerűsödik:

$$n = \frac{N \sum_{b=1}^L N_b \sigma_b^2}{(N \Delta)^2 + \sum_{b=1}^L N_b \sigma_b^2}$$

23. TÁBLÁZAT: A SZÜKSÉGES MINTAVÉTELI ARÁNYOK EV, ILLETVE RÉTEGZÉS MELLETT (%)

Állatfaj	Teljes alapsokaság		Állattartó gazdaságok	
	EV	Régiós rétegzés (arányos)	EV	Régiós rétegzés (arányos)
Szarvasmarha	3,27	3,27	3,12	3,12
Sertés	4,09	4,08	3,85	3,82
Juh	76,69	76,63	75,88	75,88

*Forrás: KSH adatbázis alapján saját számítás*

<sup>36</sup> Hajdu et al. 1994, 260. p.

**Látható (23. táblázat) az eredmények összehasonlításából, hogy bár nem tekinthető függetlennek a régió és az állatállomány nagyságának alakulása, mégis, a sztochasztikus kapcsolat gyenge intenzitása miatt a régiós rétegzés nem javítja jelentősen a mintavétel hatékonyságát, vagyis nem csökken az elérendő hibaszámokhoz tartozó mintanagyság.** Sőt, több esetben gyakorlatilag ugyanazokat az eredményeket, mintavételi arányokat kapjuk a teljes sokaságot szemlélve, illetve csak az állattartókra vonatkozóan is, mint az egyszerű véletlen kiválasztással. Ezzel az a helyzet áll elő, hogy bár használtunk (technikai) rétegzést, de a hatékonyság gyakorlatilag azonos szinten maradt. Ez annyiból meglepő, hogy szignifikáns kapcsolat van a régió és az állatállomány nagysága között, mégsem javul a hatékonyság. Pedig az arányosan rétegzett mintából becsült átlag varianciája csak szélsőséges esetben lehet legfeljebb akkora, mint az egyszerű véletlen mintából becsült átlag (ugyanakkora mintanagyság mellett). Ez abból adódik, hogy EV esetében a teljes szórás szerepet kap a mintavételi hiba számításakor, míg arányos rétegzés esetében csak a teljes szórás egy része, vagyis a belső szórás. És mivel  $\sigma^2 = \sigma_B^2 + \sigma_K^2$ , vagyis a teljes variancia (szórásnégyzet) egyenlő a külső és a belső variancia összegével, ezért a belső szórás legfeljebb akkora lehet, mint a teljes. A szélsőséges eset akkor állhat elő, ha a külső szórás nulla, vagyis teljesen független a vizsgált változó és a csoportosító ismérv. Viszont a függetlenséget korábban már elvetettük. Ezen összefüggés miatt magyarázatra szorul, hogy miért fordulhat elő hogy szinte azonos mintavételi arány szükséges az arányos rétegzésnél. A 24. táblázat és 25. táblázatban a régiós csoportosítás mellett kapott szórásfelbontás értékei láthatóak állatfajok szerint. A belső és a teljes szórás között csak századnyi különbségeket tapasztalunk. Hogyha ehhez hozzátesszük azt, hogy 7 (technikai) réteget képzünk és így alkalmazzuk az arányos rétegzést, akkor a rétegenkénti szórások kerekítésével előállhat az a furcsa helyzet, hogy bár nincsen függetlenség a változók között, de a becslés hatékonysága az EV-hez képest mégsem javul. Ez indokolja tehát a kapott eredmények alakulását.

24. TÁBLÁZAT: SZÓRÁSFELBONTÁS AZ EGYÉNI GAZDASÁGOKNÁL RÉGIÓS CSOPORTOSÍTÁS MELLETT

Szórásfelbontás	Szarvasmarha	Sertés	Juh
Belső szórás	2,48	8,54	18,80
Külső szórás	0,08	0,76	0,73
Teljes szórás	2,48	8,57	18,83

*Forrás: KSH adatbázis alapján saját számítás*



25. TÁBLÁZAT: SZÓRÁSFELBONTÁS AZ ÁLLATTARTÓ EGYÉNI GAZDASÁGOKNÁL RÉGIÓS CSOPORTOSÍTÁS MELLETT

Szórásfelbontás	Szarvasmarha	Sertés	Juh
Belső szórás	2,91	9,95	22,15
Külső szórás	0,09	0,89	0,94
Teljes szórás	2,91	9,99	22,17

*Forrás: KSH adatbázis alapján saját számítás*

Ezzel tehát a **2. hipotézist (H2)** részben számszerűen is alátámasztottuk: a régiós rétegzés (arányos allokáció mellett) nem hoz javulást a szükséges mintavételi arányban. Ez már önmagában sem túl jó hír, ráadásul további kedvezőtlen hatásként ki kell hangsúlyoznom azt a tényt is, hogy ezzel a régiók szempontjából a 7 réteg már adott, és a további rétegzési szempontok számosságát ez jelentős mértékben behatárolja, korlátozza.

26. TÁBLÁZAT: AZ EGYES ÁLLATFAJOK SZÓRÁSA, ÉS AZ EGYÉNI GAZDASÁGOK SZÁMA RÉGIÓNKÉNT

Régió	Szarvasmarha	Sertés	Juh	Gazdaságok száma
Közép-Magyarország	2,01	4,28	16,61	8 146
Közép-Dunántúl	2,40	9,86	15,26	9 016
Nyugat-Dunántúl	2,79	11,55	14,81	9 593
Dél-Dunántúl	2,46	7,52	2,48	14 147
Észak-Magyarország	1,63	3,84	14,70	13 559
Észak-Alföld	2,34	6,52	24,35	30 748
Dél-Alföld	3,02	11,66	21,18	25 529

*Forrás: KSH adatbázis alapján saját számítás*

27. TÁBLÁZAT: AZ EGYES ÁLLATFAJOK SZÓRÁSA, ÉS AZ ÁLLATTARTÓ EGYÉNI GAZDASÁGOK SZÁMA RÉGIÓNKÉNT

Régió	Szarvasmarha	Sertés	Juh	Gazdaságok száma
Közép-Magyarország	2,38	4,98	19,67	5 807
Közép-Dunántúl	2,99	12,17	19,00	5 812
Nyugat-Dunántúl	3,34	13,78	3,49	6 648
Dél-Dunántúl	2,91	8,79	17,02	10 108
Észak-Magyarország	2,00	4,65	15,76	8 928
Észak-Alföld	2,73	7,52	28,46	22 446
Dél-Alföld	3,39	13,00	23,86	20 076

*Forrás: KSH adatbázis alapján saját számítás*

Mielőtt azonban megfogalmaznánk a tézist, a **Neyman-féle optimális elosztást is meg kell vizsgálnunk**, hiszen az hozhat még kedvezőbb eredményeket, mivel itt a régiókra jellemző (helyenként ugyan csak minimálisan eltérő) szórások adják az optimalizáció alapját. Ezért ezt a lehetőséget is megvizsgálom. Tehát a becsült paraméter szórásának minimalizálására törekszik ez a módszer a mintanagyság allokálása során. **Akkor hoz jelentős javulást** – akkor számíthatunk a szükséges mintanagyság számottevő csökkenésére –, **ha az egyes állatfajok régiónkénti szórása jelentősebb eltéréseket mutat**. Másképpen fogalmazva, **ha a rétegen belül eltérő a heterogenitás szintje**. A következő táblázatok (26. táblázat és 27. táblázat) adatai ehhez nyújtanak támpontot.

A szarvasmarha esetén látjuk a leginkább „kiegyensúlyozott” szórásokat a régiók között. A sertés és a juh esetében már jelentősebb eltéréseket találunk. Ez azt jelenti, hogy a sertés és a juh esetében a Neyman-féle optimális allokáció jobban tud érvényesülni, mint a szarvasmarhánál. Nem szabad meglepődnünk azon, hogy az állattartó gazdaságokat véve alapul a szórások abszolút nagysága megnövekszik. Hiszen az átlagos, egy gazdaságra jutó állomány nagyságok megnőnek. Ugyanakkor állatfajonként a szórások nagyságrendjének változatossága hasonló a teljes alapsokaságnál tapasztaltakhoz. (A VI. számú mellékletben részletesebb tábla található a teljes sokaságra vonatkozóan. Pusztán a szórások összehasonlítása elégséges a Neyman-féle optimális allokáció hatékonyságának előzetes felméréséhez, viszont csak a szórások alapján nem lehet megítélni az egyes állatfajok régiókon belüli homogenitását. Ehhez a relatív szórások ismerete szükséges. Az itt levont következtetések tehát nem a homogenitásra vonatkoznak!)

A szükséges mintanagyság rétegenként a következő összefüggés segítségével határozható meg (az általános képletet lásd a 3.1.2 fejezetben):

$$n_h = n \frac{N_h s_h}{\sum_{h=1}^L N_h s_h}$$

Mivel a képlet az egyes rétegekre adja meg az optimumot, és a teljes mintanagyságot ( $n$ ) is felhasználja ehhez, közvetlenül nem kapjuk meg azt, hogy bizonyos hibahatár eléréséhez összességében mekkora mintára van szükségünk ilyen típusú optimalizáció mellett. Ez az allokációs technika a már kialakított mintanagyságot osztja el optimálisan a rétegek között. Ebből az következik, hogy a szükséges mintanagyság meghatározását közelítő eljárással lehet csak elvégezni. Ennek megoldására egy képletekkel „programozott” MS Excel fájlt

készítettem el. Ebben régióként és állatfajonként szerepelt minden szükséges adat (alapsokasági rétegnagyságok, szórások, előírt hibahatár) és csupán a teljes mintanagyság volt az egyetlen változtatható paraméter. Kiindulásképpen a szükséges mintaelemszámnak legalábbis a maximumát meghatározhatjuk előre. A Neyman-féle elosztás ugyanis abban a szélsőséges esetben ad az arányos elosztással megegyező hatékonyságot, ha a rétegeken belüli szórások megegyeznek. Minden más esetben (nagyon extrém alacsony elemszámoktól eltekintve) a Neyman-féle optimális rétegzés hatékonyabb lesz. Vagyis az arányos allokációhoz kiszámítható optimális mintanagyság felső korlátot ad a Neyman-féle elosztáshoz szükséges mintanagysághoz. Ezt pedig addig kellett csökkenteni, amíg az előírt hibahatárt nem teljesítette a minta (mialatt az egyes rétegekből szükséges mintaelemszám automatikusan számolódott).

28. TÁBLÁZAT: A SZÜKSÉGES MINTAVÉTELI ARÁNYOK KÜLÖNBÖZŐ MINTAVÉTELI TERVEK ESETÉN

Állatfaj	Teljes alapsokaság			Állattartó gazdaságok		
	EV	Régiós rétegzés (arányos)	Régiós rétegzés (Neyman-féle optimális)	EV	Régiós rétegzés (arányos)	Régiós rétegzés (Neyman-féle optimális)
Szarvasmarha	3,27%	3,27%	3,18%	3,12%	3,12%	3,05%
Sertés	4,09%	4,08%	3,62%	3,85%	3,82%	3,42%
Juh	76,69%	76,63%	66,44%	75,88%	75,88%	68,36%

*Forrás: KSH adatbázis alapján saját számítás*

Az eredményeket a korábbiakkal egy táblázatba (28. táblázat) foglaltam a könnyebb áttekinthetőség kedvéért. Ahogy a régiókénti szórások vizsgálatánál már „éreztük”, a szarvasmarhánál nincsen különösebb hozadéka az optimális rétegzésnek az arányoshoz képest. A másik két állatfaj esetében már látszik valamennyi javulás (csökkenés) a szükséges mintanagyságban, mind a teljes sokaság, mind az állattartó gazdaságok esetében. Ahhoz azonban, hogy a javulás mértékére is konkrétabb számokat lássunk, egy kis módszertani kitérőt kell tenni, és ezután fogalmazhatjuk meg a tézisünket a régiós rétegzéssel kapcsolatban.

### 5.2.2. Módszertani megfontolás

Amikor különböző mintavételi tervek hatékonyságát, hatásosságát szeretnénk megvizsgálni, összehasonlítani, akkor a gyakorlatban az úgynevezett tervhatásossági mutatót (DEFF – design effect) szokás használni. Ennek a mutatónak a lényege, hogy az éppen vizsgált mintavételi terv szerint az adott paraméter szórásnégyzetének becsült értékét ( $S^2$ ) hasonlítjuk ahhoz a varianciához ( $S_{EV}^2$ ), amelyet azonos nagyságú egyszerű véletlen mintából kaptuk volna meg. (Marton 1991. 32-33. o.) A mutató értéke értelemszerűen 1-nél kisebb, ha az egyszerű véletlennél hatásosabb mintavételi módszert tudunk kidolgozni. A mutató nagyságrendje is lényeges eredmény, hiszen nem mindegy, hogy mennyivel hatékonyabb, hatásosabb egyik mintavételi terv a másikhoz képest.

$$DEFF = \frac{S^2}{S_{EV}^2}$$

Természetesen nem csak az egyszerű véletlen mintavételhez képest lehet az összehasonlítást elvégezni, hanem bármilyen párosításban. (Előfeltétel ugyanakkor az azonos mintaelemszám.) Fontos hangsúlyozni azt is, hogy a DEFF mutató értéke többcélú összeírások esetében változónként (jelentősen) eltérő is lehet. Vagyis ez nem egy összefoglaló mutató a mintavételi terv hatásosságának jellemzésére, hanem minden egyes vizsgált paraméterre külön számszerűsítendő és értékelendő.

Elméleti szempontból nagy jelentősége van ennek a mutatónak, hiszen így tudjuk közvetlenül megítélni általánosságban, hogy a különböző mintavételi módszerek hatékonyságukat tekintve hogyan viszonyulnak egymáshoz. A gyakorlatban egy esetleg bonyolultabb, idő- és költségigényesebb mintavételi terv minőségjavító-hatását láthatjuk számszerűen. Ez segíthet abban, hogy el tudjuk dönteni, megéri-e más, esetleg összetettebb módszert választani.

A gyakorlati munka során azonban a probléma, a mintavételi tervek összehasonlításának igénye gyakran nem ebben a formában merül fel. (Lásd H4-t a bevezetőben.) Ahogy a dolgozat példájában is, nem ritka, hogy előírás a mintavételi hiba nagysága, illetve aránya (relatív mintavételi hiba). Ekkor a mintavételi tervek hatásosságának összehasonlítása során sok értelme nincsen a mintavételi hibán, a vizsgált paraméter becsült szórásnégyzetén keresztül vizsgáldni, és minden módszerre, ugyanolyan mintanagyság mellett elvégezni a

szükséges számításokat. Sokkal inkább célravezető és gyakorlatiasabb megoldásnak tűnik a mintaelemszámok összehasonlítása az előírt hibanagyságok teljesítése mellett. Ekkor tehát minden mintavételi terv esetén (közelítőleg) kiszámítjuk az adott hibahatár teljesítéséhez szükséges mintanagyságot. Ráadásul ezekre az eredményekre mindenképpen szükségünk lesz, hiszen végül bármelyik megoldást is választjuk, a pontossági előírásnak megfelelő mintával kell majd a továbbiakban dolgoznunk. Az így kapott értékeket is viszonyíthatjuk egymáshoz. Ekkor is megtehetjük azt, hogy kiindulásnak minden tervet az egyszerű véletlen mellett szükséges elemszámhoz viszonyítunk, de a különböző módszerek közvetlen egymáshoz hasonlítása is lehetséges. Az arányosan rétegzett és az egyszerű véletlen mintavétel összehasonlítása a következő képlettel írható le:

$$SEFF_{(AR/EV)} = \frac{n_{(AR)}}{n_{(EV)}}$$

A mutatót mérethatásosságnak<sup>37</sup> lehetne elnevezni (size efficiency – SEFF), hiszen a vizsgálódások háttérében nem a mintanagyság azonos (fixált), hanem a paraméter becsült varianciája, és a mintanagyságokat, a különböző mintavételi tervhez tartozó minta méretét hasonlítjuk össze. Sikerült tehát egy olyan mutatót kialakítani, ami a dolgozat magját képező mintavételi probléma vizsgálatához jobban igazodik. (H4)

## 2. tézis

**Az úgynevezett mérethatásosság mutató (SEFF) a gyakorlati munka során a tervhatásossági mutatónál (DEFF) jobban alkalmazható mérőszám a különböző mintavételi tervek összehasonlítására, hiszen ekkor adott standard hiba mellett hasonlítjuk össze a szükséges mintanagyságokat.**

(Ennek mintájára az is elképzelhető lenne, hogy nem a mintanagyságokat, hanem az összeírás lebonyolításához szükséges költségeket hasonlítjuk össze. A költségek kalkulálása

---

<sup>37</sup> Lehetséges, hogy nem ez a legjobb elnevezés. Felmerült még a nagysághatásosság, fordított tervhatásosság is.

azonban a tényleges minta kiválasztása előtt csak közelítve határozható meg, és így is nagyon idő- és munkaigényes folyamat lenne minden mintavételi terv esetére kiszámítani.)

Visszatérve a régiós rétegzés és az egyszerű véletlen mintavétel összehasonlításához egy ilyen mutató felhasználása a következő eredményeket adja:

29. TÁBLÁZAT: KÜLÖNBÖZŐ RÉGIÓS RÉTEGZÉS ÉS AZ EV MINTAVÉTELI TERV ÖSSZEHASONLÍTÁSA

Állatfaj	Teljes sokaság			Állattartó gazdaságok		
	SEFF <sub>(AR/EV)</sub>	SEFF <sub>(NO/EV)</sub>	SEFF <sub>(NO/AR)</sub>	SEFF <sub>(AR/EV)</sub>	SEFF <sub>(NO/EV)</sub>	SEFF <sub>(NO/AR)</sub>
Szarvasmarha	1,001	0,972	0,971	1,001	0,978	0,976
Sertés	0,994	0,882	0,887	0,992	0,888	0,895
Juh	0,997	0,865	0,867	1,000	0,901	0,901

*Forrás: KSH adatbázis alapján saját számítás*

A 29. táblázatban a mutató értékei azt mutatják, hogy az arányos rétegzés két esetben minimálisan még ront is az egyszerű véletlen kiválasztáshoz képest, a korábban bemutatott okok miatt.

Ezzel szemben, a teljes sokaságot véve alapul, a Neyman-féle optimális allokáció mellett a szükséges mintanagyság néhány százalékkal lecsökken. A juhoknál tapasztaljuk a legtöbb hozadékot (9,9-13,3 százalékos mintanagyság csökkenést). Amikor arról döntünk, hogy az optimális elosztás jelentősen javítja-e a mintavétel hatásosságát, akkor azt is szem előtt kell tartanunk, hogy az egyes változók szórásait csak becsléssel tudjuk közelíteni, ami plusz munkát igényel, többlet alapadat-szükséglete van, és nem is ad teljesen pontos értékeket. Éppen ezért nem biztos, hogy ilyen viszonylag kismértékű javulás reménye elegendő ok a nagyobb munka elvégzéséhez.

A teljesség kedvéért a harmadik, leggyakrabban említett allokációs módszert, az egyenletes elosztást is megvizsgáltam az EV kiválasztáshoz képest. A mérethatásossági mutató értéke teljes sokaságnál szarvasmarhára 1,33, sertésre pedig 1,32; állattartó gazdaságoknál szarvasmarhára 1,41, míg sertésre 1,91. A juhoknál nem is lehet egyenletesen elosztani a szükséges mintaelemszámot, mert annyira nagy mintára van szükség, hogy azt egyenlő arányban felosztva néhány régió esetében nagyobb mintaelemszám jönne ki eredményül, mint a régió teljes gazdaságszáma. Az egyenletes rétegzés tehát biztosan nem adja meg a jó irányt, a hatékony megoldást, ezért részletesen azt nem is mutatom be.

Az eredmények alapján a 2. hipotézis feltevése csak részben igazolódott. A tézisnek ezért árnyaltabb megfogalmazást kell adni a H2-höz képest:

### 3. tézis

**A régió, mint rétegeképző ismerv csak a Neyman-féle optimális allokáció felhasználásával hoz javulást az egyszerű véletlen mintavételhez képest, és a javulás mértéke ugyan állatfajonként különböző, de sehol sem nevezhető jelentősnek.**

A számítások bemutatásakor már eddig is kitértem a teljes sokaságot, illetve az állattartó gazdaságokat alapnak tekintő szemlélet különbségeire. A bevezetőben megfogalmazott **1. hipotézist (H1)**, amely szerint hatékonyabb csak az állattartó gazdaságokat alapsokaságnak tekintve kidolgozni a mintavételi tervet, **nem támasztják alá az eddigi eredmények**. A mintavételi arányok nem mutatnak lényegesen alacsonyabb szintet. (Természetesen a szükséges minta nagysága még így is kisebb, hiszen 27 százalékkal kisebb számosságú az állattartók sokasága a teljeshez képest. Ez azonban a hatékonyságot módszertani szempontból nem jellemezheti.) Ahhoz azonban, hogy a H1 elfogadásáról dönteni tudjunk, szükséges a további mintavételi lehetőségek elemzése is.

#### 5.2.3. Rétegzés nagyságkategóriák szerint

A régió tehát egy nem igazán hatékony, de mégis kötelező rétegzési szempontot jelent. Emellé azonban mindenképpen szükségesnek látom olyan rétegeképző ismerv(ek) keresését, ami szoros sztochasztikus kapcsolatban áll a vizsgált változókkal, és így a kapcsolati rétegzés is megvalósítható lenne, ami már jelentősen javíthatná a mintavételi terv hatásosságát.

Korábbi tapasztalataim azt mutatják, hogy a mezőgazdaságban is „nagyjából” érvényesül a 80-20 százalékos arányeltolódás. Legalábbis annyiban, hogy a gazdaságok viszonylag kis arányánál található az állomány nagyobb része. Ez összeírási, mintavételi szempontból azt jelenti, hogy az értékösszeg viszonylag nagy része „megfogható” egy szűkebb sokaság megfigyelésével. Az értékösszeg kisebb szelete az, amit egy valóban nagy számosságú sokaság összeírásával kell megbecsülni. Ugyanakkor ennek a nagyobb sokaságnak a becslése

során keletkező hiba összességében kisebb „jelentőségű” az alacsonyabb értékösszeg miatt. Ezt a 30. táblázatban foglalt eredmények is alátámasztják, megközelítőleg. Ez a tény pedig azt vetíti előre, hogy célszerű lehet úgy tekinteni a „nagyobb” és „kisebb” gazdaságok csoportjaira, ahogyan a gazdasági szervezetek és az egyéni gazdaságok vonatkozásában. Hiszen a „nagyobb” egyéni gazdaságok feltételezhetően hasonlóan jól összeírhatóak lennének összeírók nélkül is, feltéve, hogy naprakész és pontos azonosító adataikat (név, cím, telefonszám, email-cím) ismerjük.

Emellett az is jellemző, hogy néhány gazdaság kiugróan magas állománnyal rendelkezik, ami még a „nagyok” között is kiemelkedőnek számít, és ennek megfelelően a szórások magasak maradnak a „nagyok” csoportjában. Az állattartó gazdaságok koncentrációját már korábban, a 4.2 fejezetben is jellemeztem a Lorenz görbék segítségével (16. ábra), ami ugyanezt támasztja alá.

30. TÁBLÁZAT: A „KISEBB” GAZDASÁGOK ARÁNYA, ÉS A HOZZÁJUK TARTOZÓ ÁLLATÁLLOMÁNYI ARÁNYOK, 2007

Állatfaj	„Kisebb” gazdaságok aránya (%)	„Kisebb” gazdaságok állomány aránya (%)
Szarvasmarha	79	28
Sertés	75	27
Juh	82	15

*Forrás: Saját szerkesztés.*

A legjobb rétegeképző ismérv maga a vizsgált változó lenne, mivel azonban erre szeretnénk becslést kapni a mintavételes összeírásból, ezért ez nyilván nem áll rendelkezésünkre. A legjobban korreláló változó azonban lehet a vizsgált ismérv egy korábbi (nem túl régi) összeírásból származó értéke, illetve eloszlása. A mezőgazdaság statisztikában a 10 évenkénti censzusok nagyon jó alapot adnak, legalábbis az ÁMÖ utáni pár évben. A távolabbi időszakban pedig a 3-4 évente közbeiktatott gazdaságszerkezeti összeírások (amelyek egy rendszeres éves összeírásnál jóval nagyobb mintával dolgoznak) nyújthatnak hasznosítható eredményeket, támpontot. Tehát adottságként számíthatunk arra, hogy rendelkezünk néhány évnél nem régebbi adatokkal a sokaság egy viszonylag nagy részére a vizsgált változóról. Ugyanezért felmerül a hányadosbecslés hatékonyságjavító tulajdonsága is, hiszen ha vannak viszonylag erősen korreláló adataink korábbi időszakról, akkor az aktuális/régebbi adatok hányadosának segítségével pontosabb becslések készíthetők. (McCallion 1992) Igen ám, de ehhez arra is szükség volna, hogy a mintába kerülő összes gazdaságról rendelkezésre álljon a



korábbi időszak megfelelő adata. Ez azonban csak a censzusokból van meg, vagyis a census utáni néhány évben nyílik csak lehetőség a hányadosbecslés felhasználására. A hányadosbecslés nyújtotta előnyöket ki is használták a KSH-ban a 2000-es ÁMÖ utáni néhány évben, de hosszabb távon, illetve folyamatosan erre nincsen lehetőség.

A különböző állatfajok esetében nyilvánvalóan homogénebb csoportokat tudunk létrehozni, ha nagyságkategóriákat alakítunk ki. Mivel 3 állatfajt vizsgálunk, ezért a fajonkénti 2 csoport kialakítása mellett döntöttem. Ha ennél több kategóriát választunk, akkor a rétegek száma túlságosan megnő, ami a későbbi régiós kombinálás eredményeképpen kezelhetetlen számú réteg kialakulását hozhatja magával. A másik oldalról pedig az is problémát jelent, hogy a túl sok részsokaságra darabolás eredményeképpen az egyes „cellákban” szélsőséges esetben nem marad megfelelő számú gazdaság.

Így tehát minden állatfaj esetében lesz egy „nagyobb”, és egy „kisebb” gazdaságokat tartalmazó réteg. A kérdés az, hogy hol lesz az optimális réteghatár. Azonban ez a feltétel, hogy „optimális réteghatár rétegzett mintavétel mellett” pontosításra szorul, mert a végeredmények szempontjából egyáltalán nem mindegy, hogy milyen allokációs módszer társul a rétegzés mellé, és melyiknél keressük a legjobb megoldást adó réteghatárt. Ennek megfelelően sorra veszem a háromféle elosztási technikát, bemutatva és összehasonlítva az optimumokat és azok különbségeit.

Az optimális réteghatár elnevezést a későbbiekben is fogom használni, ezért már itt szeretném összefoglalni, hogy mit is értek ez alatt az állatállomány becslése szempontjából. **Optimális réteghatár** (OPS – optimum point of stratification) az az állomány létszám lesz, amely úgy osztja két rétegre a gazdaságokat (kisebbekre és nagyobbakra), hogy azzal a becslés standard hibája a legalacsonyabb lesz (azonos mintaelemszámot feltételezve). Természetesen ez függ az allokációs módszertől, és szélsőséges eseteken kívül az optimum nem esik egybe a mintanagyság különböző elosztásai mellett. Állatállomány számbavételről lévén szó, az optimális réteghatár mindig egész szám lesz. Az azonban nem biztos, hogy egy olyan pont van, ami eleget tesz a fenti feltételnek. Elméletileg az is előfordulhat, hogy több (egymás melletti) érték mellett is ugyanolyan minimális hibanagyságú becslés érhető el.

## Arányos allokáció

Az arányos allokáció, ahogy a 3.1.2 fejezetben részletesen bemutatam, azt jelenti, hogy a teljes mintanagyságot a rétegek nagyságával arányosan osztjuk szét a rétegek között.

Arányos elosztás mellett elméletileg ott volna a legcélszerűbb meghúzni a határvonalat, ahol a rétegekbe osztott sokaság a legkisebb belső szórást produkálja. Ennek vizsgálata céljából egy egyszerű iterációs eljárást alkalmaztam. Minden állatfaj esetében önkényes, egyre növekvő réteghatárokat választottam, és kiszámítottam a különböző szempontú alapsokaságok mellett érvényesülő belső szórások nagyságát. Az összefoglaló eredményeket a VII. számú melléklet tartalmazza. Minden esetben érzékelhető, hogy a belső szórás nagysága egy bizonyos pontig csökkenő tendenciát mutat, majd a minimumát elérve növekedni kezd. A legkisebb belső szórásokhoz tartozó réteghatárok mutatják meg, hogy hol célszerű meghúzni a határvonalat a két réteg között, állatfajonként. A belső szórás minimumát, vagyis a szóráshányados maximumát jelentő réteghatár meghatározására rendelkezésünkre áll egy másik módszer is. Amennyiben az adatokat egy változó szerint 2 csoportba soroljuk K-közepű klaszteranalízis segítségével, akkor pontosan a keresett osztópont fog kialakulni.

Mindegy, hogy milyen szemléletben vizsgáljuk a sokaságot (teljes sokaság, csak az állattartók sokasága), az a talán speciálisnak mondható eset áll fenn, hogy az optimum ilyen értelemben ugyanott helyezkedik el. Ennek az az oka, hogy a réteghatár feletti állattartók létszáma, szórásai mindkét szemléletben ugyanazt mutatják, és a belső szórást markánsan ez determinálja. Csak a réteghatár alatti kisebb gazdaságok száma és összetétele módosul. A teljes sokaság esetében sok nullás értékkel rendelkező gazdaság csapódik a kisebbek rétegéhez, de ez jelentősen nem módosítja a belső szórás nagyságát. A 31. táblázat tartalmazza az eredményeket. A szükséges mintanagyságokat a fentiekben bemutatott módszerrel számítottam ki. Az optimumokat színesen kiemeltem a táblázatban, mert mellettük néhány további réteghatár mellett is kigyűjtöttem a mintavételi arányokat, az összehasonlíthatóság kedvéért.

31. TÁBLÁZAT: OPTIMÁLIS RÉTEGHATÁR ARÁNYOS ALLOKÁCIÓ MELLETT

Állatfaj	Réteghatár (állatállomány nagyság)	Teljes sokaság		Állattartók	
		n	n/N (%)	n	n/N (%)
Szarvasmarha	5	1 675	1,51	1 153	1,44
	15	1 198	1,08	846	1,06
	30	1 465	1,32	1 001	1,25
Sertés	60	2 363	2,13	1 581	1,98
	100	2 263	2,04	1 445	1,81
	140	2 297	2,07	1 528	1,91
Juh	140	52 479	47,3	36 997	46,3
	180	51 813	46,7	36 517	45,7
	220	53 078	47,8	37 363	46,8

*Forrás: KSH adatbázis alapján saját számítás*

Az állattartókra korlátozódó sokasági szemlélet érdemi különbségeket nem mutat a szükséges mintavételi arányok tekintetében, a vizsgált sokaságon belül. Nagyság szerinti arányos rétegzés esetén tehát nem igazolható az eredeti feltevés a szűkebb szemléletű alapsokaság hatékonyságáról.

A 31. táblázatban szereplő, nem optimum pontok melletti eredmények még egy érdekes tulajdonsága az, hogy az optimális réteghatártól távolabbi értékekhez csak néhány tized százalékkal magasabb mintavételi arányok tartoznak. Vagyis, ha nem is sikerül előzetes adatokból, korábbi összeírásokból teljesen pontos szórásokkal számolnunk, akkor sincsen nagy baj, hiszen nagy különbségek nincsenek az arányos elosztásnál kiszámított optimális réteghatár környezetében. Ez mindenképpen megnyugtató és jó tulajdonság.

Talán ránézésre is látszik, hogy a sokaság két részre osztása – a legszorosabb sztochasztikus kapcsolatot mutató réteghatár mentén – jelentős javulást okoz az egyszerű véletlen mintavételhez képest. A konkrét összehasonlítást a mérethatásossági mutató segítségével a 32. táblázat tartalmazza.

32. TÁBLÁZAT: SEFF ARÁNYOS NAGYSÁG SZERINTI RÉTEGZÉS ÉS EV MINTAVÉTEL MELLETT

Állatfaj	SEFF(AR/EV)		Szóráshányados (H)
	Teljes sokaság	Állattartók	
Szarvasmarha	0,327	0,467	0,91
Sertés	0,510	0,651	0,85
Juh	0,611	0,854	0,93

*Forrás: KSH adatbázis alapján saját számítás*

A legjelentősebb javulást a szarvasmarhákánál látjuk, ahol kevesebb, mint fele lett a szükséges mintaelemszám. A sertéseknél is 35-50 százalékkal sikerül így lecsökkenteni a mintanagyságot. Az amúgy is kiugróan magas mintavételi arányt igénylő juhok esetében mutatkozik a legalacsonyabb mértékű javulás. A táblázatban a szóráshányados mutatóját is feltűntettem, hogy lássuk, milyen szoros sztochasztikus kapcsolat mellett alakultak ki az optimumok. Mindhárom fajra elmondható, hogy erős, szoros sztochasztikus kapcsolat áll fenn az állatállomány nagysága és a kialakított 2 réteg között. Az egyszerű véletlenhez képest nyert hozadék viszont nincsen arányban a kapcsolatszorossági mérőszám értékével. A juhoknál találjuk a legszorosabb kapcsolatot, míg itt hozott a legkisebb javulást az arányos rétegzés. Ez első ránézésre talán ellentmond a „megérzésnek”, miszerint minél szorosabb a kapcsolat a rétegeképítő ismerv és a vizsgált változó között, annál hatékonyabb lesz a rétegzett mintavétel. Az elért eredmények tükrében viszont mélyebben is belegondolva a problémába, a „szorosabb kapcsolat – kisebb hatékonyság javulás” oka valószínűleg a változó terjedelmének alakulásában és az ebből adódó, abszolút értelemben vett nagyobb szórásértékekben keresendő. A példánkban ugyanis a juhoknál találkozunk a legnagyobb terjedelemmel.

Arányos rétegzés mellett az optimum ilyen módon történő meghatározása meglehetősen bonyolult, különösen akkor, ha nem áll rendelkezésre frissnek tekinthető adatbázis a gazdaságok jellemzőiről, történetesen a vizsgált változók egyedi értékeiről. Ezek hiányában ugyanis a belső szórás nem számszerűsíthető.

Dalenius (1950) bebizonyította, hogy az optimális réteghatárhoz jutunk el, ha a következő feltétel teljesül:

$$x_k = \frac{1}{2}(\bar{x}_{k-} + \bar{x}_{k+})$$

ahol

$x_k$  – az optimális réteghatár (OPS)

$\bar{x}_{k-}$  – a réteghatár alatti elemek átlagos értéke

$\bar{x}_{k+}$  – a réteghatár feletti elemek átlagos értéke

Az összefüggés tehát az optimális réteghatár esetén fog teljesülni. Ez sem egy közvetlen megoldás, hiszen egy lépésben nem jutunk el a végeredményhez. Első lépésben önkényesen

választunk egy réteghatárt, kiszámítjuk a feltételben szereplő tagokat. A következő lépésben módosítjuk a réteghatár értékét (növeljük vagy csökkenjük), és így is elvégezzük a számításokat. A lépéseket addig ismételjük, amíg az egyenlőség teljesül. A számítások elvégzése egyszerűbb, mint a belső szórás felhasználásakor használt módszernél, viszont itt is szükséges az egyedi adatok ismerete ahhoz, hogy a réteghatárt mozgathassuk, és a változások mellett az átlagokat számszerűsíteni tudjuk.

Bár Dalenius közvetlenül nem mondja ki, de a gyakorlatban a belső szórás minimumából és a fentiekben bemutatott összefüggésből adódó optimum ugyanaz (több optimum nem is lehetséges).

### Neyman-féle optimális allokáció

A Neyman-féle optimális allokációt már korábban is ismertettem. A nagyság szerinti rétegzéshez ugyanazt az eljárást használtam fel, amelyet a régiós rétegzésnél a 5.2.1 fejezetben már bemutatam. Képletekkel „programozott” MS Excel fájl segítségével, iterációs eljárással jutottam el az optimális réteghatárokhoz. Annyival volt egyszerűbb a helyzet, hogy csak 2 réteget vizsgáltam, míg a régióknál ez 7 réteget jelentett.

33. TÁBLÁZAT: SZÜKSÉGES MINTANAGYSÁG KÜLÖNBÖZŐ RÉTEGHATÁROK MELLETT (NEYMAN-FÉLE OPTIMÁLIS)

Állatfaj	Réteghatár (állomány nagyság)	Teljes sokaság		Állattartók	
		n	n/N (%)	n	n/N (%)
Szarvasmarha	30	1 287	1,16	886	1,11
	20	810	0,73	559	0,7
	15	588	0,53	423	0,53
	10	388	0,35	279	0,35
	5	200	0,18	152	0,19
	2	111	0,1	96	0,12
Sertés	20	721	0,65	535	0,67
	10	577	0,52	463	0,58
	6	566	0,51	495	0,62
	5	566	0,51	519	0,65
	4	588	0,53	567	0,71
	3	644	0,58	647	0,81
Juh	20	-	-	-	-
	15	-	-	-	-
	10	-	-	1 820	2,28
	8	1 997	1,8	1 852	2,32
	6	2 030	1,83	1 916	2,4

Forrás: KSH adatbázis alapján saját számítás

Az arányos elosztáshoz képest nagyon eltérőek az optimális réteghatárok, jóval alacsonyabb szinten állnak. (33. táblázat)

2 szarvasmarhánál meghúзва a határt kapjuk elméletileg a legkedvezőbb mintavételi arányt. De ennél az állatfajnál szinte minden vizsgált réteghatár 1 százaléknál alacsonyabb kiválasztási arányt hozott eredményül, ami rendkívül ígéretesnek mondható eredmény a gyakorlat számára. Ez azért is nagyon jó hír, mert ezzel azt is kimondtam, hogy a szórások esetleges pontatlanabb ismerete mellett sem tudunk „nagyot hibázni”, hiszen 30-as állatállomány alatt szinte bárhová húzva a réteghatárt meglehetősen alacsony mintanagyságra van csak szükségünk hatásos becslés eléréséhez.

Sertéseknél 5-10 állat között húzódik az optimum. Ennél pontosabban nem is érdemes meghatározni, mert az elemszámokból látható, hogy szinte jelentéktelen különbség van az értékek között. A szarvasmarhához hasonlóan itt is jó eséllyel érünk el hatékony rétegzést a réteghatárok egy viszonylag szélesebb tartományán keresztül. A szükséges mintavételi arány fél százalék körüli.

A juhoknál a kihúzott cellákba azért nem kerültek értékek, mert a „nagyok” rétegéből a Neyman-féle optimális allokáció elveinek megfelelően nagyobb mintaelemszám lett volna indokolt, mint maga a rétegnagyság. A gyakorlatban ez tehát azt jelenti, hogy minden elem bekerülne a „nagyobb” gazdaságok rétegéből, vagyis a kiválasztási arány itt 100 százalékos lenne. Mivel ezt a módszert (take all – take some) a későbbiekben tárgyalom, itt nincsen jelentősége a számoknak (hiszen igazából nem tükrözik a Neyman-féle optimumot). A másik két állatfajhoz képest a juhoknál találjuk a legmagasabb értékeket. De még így is – a korábbiakban bemutatott jellemzők tükrében – nagyon ígéretesek a táblázatban olvasható, viszonylag alacsony kiválasztási arányok.

Mindezek fényében, a bevezetőben felvetett 6. hipotézist, miszerint a kevésbé aszimmetrikus eloszlással rendelkező szarvasmarha és sertés esetében „egyszerűbb” lesz viszonylag alacsony mintanagyságot eredményező mintavételi tervet kialakítani, sikerült igazolni. Megfogalmazva a tézist:

#### 4. tézis

**Kisebb terjedelemmel rendelkező eloszlások esetében a nagyságkategóriák szerinti kapcsolati rétegzésnél a minimálisan szükségeset jelentősen nem meghaladó mintaelemszámot eredményező mintavételi tervekhez tartozó réteghatár nem egy pontba, hanem egy nagyobb tartományba esik, ami megkönnyíti a meghatározását.**

Triviálisnak tűnhet, hogy **mindhárom állatfajnál a szűkebb szemléletű alapsokaság mellett találjuk a legalacsonyabb szükséges mintaelemszámot, hiszen eleve lényegesen kisebb sokaságból indulunk ki.** Ugyanakkor fontos észrevenni azt a tényt, hogy **a mintavételi arány minden esetben nagyobb az állattartó gazdaságoknál.** Ez azt jelenti, hogy az 1. hipotézist nem sikerült bizonyítani. Hiába szűkítjük a sokasági szemléletet, a szükséges mintavételi arány nem lesz alacsonyabb. Vagyis az erre vonatkozó 1. hipotézist ezek alapján el kell vetni. Tézisként ugyanakkor megfogalmazható az ellentéte:

#### 5. tézis

**Nem eredményez alacsonyabb kiválasztási arányokat az alapsokasági szemlélet szűkítése, és csak az állattartó gazdaságok figyelembe vétele.**

Ezek után, még az is megfogalmazódott bennem, hogy nem feltétlenül éri meg szűkíteni a sokasági szemléletet, és ezzel bonyolítani a minta kijelölését (illetve függetleníteni a mintát a föld- és vetésterületi összeírásoktól), mert módszertani szempontból nem nyerünk eleget ezzel.

Az optimális réteghatárokhoz rendelhető mérethatásossági mutató értékei (34. táblázat) azt mutatják, hogy mind az egyszerű véletlen, mind az arányos rétegzéshez képest a szórásokat szem előtt tartó optimalizáló allokációs eljárás jelentősen csökkent a szükséges mintanagyságot. A szarvasmarha esetében 82-97 százalék közötti a csökkentés mértéke. Sertéseknél 68-81 százalék közé esik a javulás. Ez annak fényében még inkább jó eredménynek tekinthető, hogy már eddig is elfogadható mértékű kiválasztási arányokat találtunk a korábban bemutatott mintavételi módszerekkel erre a két állatfajra. Leginkább

szembetűnő változás mégis a juhoknál érzékelhető, ahol 95 százaléknál is többet csökkent a szükséges mintaelemszám. Hozzá kell tenni, hogy ez volt az az állatfaj, ahol egyébként az eddigi módszerekkel csak extrém nagyarányú mintával tudtuk teljesíteni a pontossági követelményeket. Ugyanakkor, az eddig „kezelhetetlennek” tűnő tulajdonságokkal rendelkező eloszlás ezek szerint mégiscsak „betörhető” a megfelelő rétegzéssel és a Neyman-féle optimális allokáció alkalmazásával.

34. TÁBLÁZAT: SEFF KÜLÖNBÖZŐ RÉTEGHATÁROK MELLETT<sup>38</sup>

Állatfaj	Réteghatár (állomány nagyság)	SEFF (NO/EV)		SEFF(NO/AR)	
		Teljes sokaság	Állattartók	Teljes sokaság	Állattartók
Szarvasmarha	30	0,352	0,489	1,074	1,047
	20	0,221	0,308	0,676	0,661
	15	0,161	0,233	0,491	0,500
	10	0,106	0,154	0,324	0,330
	5	0,055	0,084	0,167	0,180
	2	0,030	0,053	0,093	0,180
Sertés	20	0,162	0,241	0,319	0,370
	10	0,130	0,209	0,255	0,320
	6	0,128	0,223	0,250	0,343
	5	0,128	0,234	0,250	0,359
	4	0,132	0,255	0,260	0,392
	3	0,145	0,292	0,285	0,448
Juh	20	-	-	-	-
	15	-	-	-	-
	10	-	0,042	-	0,049
	8	0,024	0,042	0,039	0,050
	6	0,024	0,044	0,039	0,051

*Forrás: KSH adatbázis alapján saját számítás*

A különböző allokációs technikák egymáshoz viszonyított hatékonyságát 2 réteg esetén általánosan is megvizsgáltam már a két réteg egymáshoz viszonyított szórásainak és elemszámainak arányában. A már korábban is említett „A rétegzett mintavételről” című tanulmány a Statisztikai Szemlében jelent meg 2011 szeptemberében.

A rétegzett mintavétel általános bemutatásánál (3.1.2 fejezet) azt emeltem ki, hogy a kapcsolati rétegzés esetén törekszünk a vizsgált paraméterrel leginkább szoros sztochasztikus

<sup>38</sup> NO = Neyman-féle optimális elosztás



kapcsolatban álló rétegeképző ismérvet választani. Láthattuk az arányos rétegzésnél az optimum valóban a legszorosabb kapcsolat mellett állt fenn. Ehhez képeset a Neyman-féle optimális elosztáshoz tartozó OPS-ek mellett a 35. táblázatban szereplő értékeket találjuk a kapcsolat szorosságára vonatkozóan.

35. TÁBLÁZAT: SZÓRÁSHÁNYADOS ÉRTÉKEK KÜLÖNBÖZŐ RÉTEGHATÁROK MELLETT

Állatfaj	Réteghatár	Szóráshányados
Szarvasmarha	15*	0,91
	2	0,81
Sertés	100*	0,85
	10	0,75
	6	0,71
	5	0,70
Juh	180*	0,93
	10	0,75
	8	0,73

*\*Az arányos rétegzés mellett érvényes optimumok.*

*Forrás: KSH adatbázis alapján saját számítás*

Mivel korábban megfogalmaztuk, hogy az arányos rétegzés optimuma a rétegeképző ismérv és a vizsgált változók között fennálló legszorosabb kapcsolat (legmagasabb szóráshányados érték) mellett valósul meg, ezért természetszerű, hogy a Neyman-féle optimális elosztás optimuma ennél csak gyengébb intenzitású kapcsolatot mutathat (illetve szélsőséges esetben legfeljebb éppen olyan szorosat). A kérdés az, hogy mennyivel gyengébbet. A 32. táblázatra tekintve azt a következtetést vonhatjuk le, hogy bár a kapcsolat erőssége alacsonyabb mértékű, minden esetben szoros marad, vagyis 0,7 vagy afeletti.

### **Take all – take some**

A KSH-ban töltött évek tapasztalata azt mutatja, hogy célszerű úgy kialakítani a „nagyok” csoportját, hogy a „take all” elvet érvényesíteni lehessen, vagyis a nagyobb gazdaságok teljes körűen bekerülhessenek a mintába. Ezzel a szórások, és így a hibaszámok tekintetében jelentős nyereséget realizálhatunk, amellett, hogy a nagyobb méretű gazdaságok megfigyelése a 3.2 fejezetben bemutatottaknak megfelelően történhet postai úton is, ami az összeírás költségeit és szervezésigényét jelentősen lecsökkenti. Ezért felmerült bennem az az elképzelés

is, hogy optimális réteghatárt határozzak meg egy ilyen speciális elosztás mellett is. Az eredményeket pedig a többi allokációs technika hatékonyságának fényében fogom értékelni.

**A jelentős (jobboldali) aszimmetriát mutató (erősen pozitív ferdeségű) eloszlások esetére** Hidioglou (1986) tárgyal egy olyan módszert, ahol **a nagyobb méretű egyedeket teljes körűen kell megfigyelni, és csak a küszöbérték alatti rétegben kerül sor tényleges mintavételre.** Ennek a „take all – take some” (TATS) elvnek a háttérében az húzódik meg, hogy a nagyobb egyedekre ezáltal „hibátlan” eredményeket kapunk, ráadásul ezek összességében a teljes értékösszeg jelentős szeletét adják. Emellett pedig a magas értékekkel rendelkező egységek „lenyesésével” a maradék réteg is homogénebb lesz az alapsokasághoz képest, vagyis csökkennek a szóródások. A gyakorlatban ilyen aszimmetrikus eloszlásra jó példát kínál a kiskereskedelem, ahol néhány, nagyon magas piaci részesedéssel rendelkező lánc adja a forgalom jelentős hányadát, és sok, viszonylag kis egység a többit. A mezőgazdaság is ilyen jellemzőkkel bír, ahogy azt már korábban is bemutattam, bár némileg eltérő szemléletben. A kiskereskedelemben joggal feltételezhetjük, hogy egy nagy forgalommal rendelkező egység vagy gazdasági szervezet minden más paraméterét tekintve is nagynak tekinthető. A mezőgazdaságban ugyanakkor attól, hogy valamilyen értékalapú szemléletben (bruttó termelési érték vagy standard fedezeti hozzájárulás) nagynak minősül egy gazdaság, még nem jelenti azt, hogy minden fontos változó (állatfajok és földterület) vonatkozásában is nagy lesz. Akár 1-1 állatfajra, vagy 1-1 művelési ágra, növényfajra specializálódó gazdaság is lehet nagy, mialatt a legtöbb paramétere egyébként nagyon alacsony, nem ritkán nulla. Ezért a „nagyok” besorolását sokkal célravezetőbb változónként meghatározni, ahogy ezt a következőkben alkalmazom is.

Azt is szeretném kiemelni, hogy a TATS módszer végeredményben a Neyman-féle optimális elosztás egy speciális válfajának is tekinthető. Hiszen, amennyiben a Neyman-féle optimális elosztás bizonyos, magas szórással bíró rétegből az alapsokasági elemszámnál ( $N_h$ ) nagyobb mértékű mintát ( $n_h$ ) követelne meg, akkor maximálisan kihasználva az alapsokasági rétegnagyságot TATS elvű minta jön létre. Ez persze nem feltétlenül jelenti azt, hogy ez egy optimális megoldás is lesz egyben. Ez viszont azt sugallja számomra, hogy olyan osztópont, réteghatár mellett lehet a TATS az optimális megoldás, ahol a Neyman-féle allokáció kritériumának már nem tudunk megfelelni a  $n_h > N_h$  miatt.

Fel tudjuk írni azt a speciális esetet, amikor a TATS szerinti minta pontosan megegyezik a Neyman-féle optimális elosztásával, két réteget feltételezve. A korábban bemutatott képletet

felhasználva, feltételezve, hogy az első réteget teljes körűen megfigyeljük a következőt kapjuk:

$$n_1 = \frac{N_1 \sigma_1}{N_1 \sigma_1 + N_2 \sigma_2} \times n$$

mivel feltettük, hogy  $n_1 = N_1$ , ezért

$$n \sigma_1 = N_1 \sigma_1 + N_2 \sigma_2$$

$$(n - N_1) \sigma_1 = N_2 \sigma_2$$

és mivel  $N_1 = n_1$ , ezért  $(n - N_1)$  helyébe  $n_2$ -t írva

$$n_2 = N_2 \frac{\sigma_2}{\sigma_1}$$

Vagyis abban az esetben egyezik meg a két módszer által meghatározott megosztás, ha a „take some” rétegből éppen a két réteg szórásainak arányában választunk mintát. Azt is kiolvashatjuk a képletből, hogy ez a kiválasztási arány ráadásul meglehetősen alacsony lesz, hiszen az első réteghez tartozó szórás jóval magasabb (azért is figyelünk meg abból minden elemet), mint a másodikban, vagyis a hányadosuk nagyon kicsi lesz. Hangsúlyozandó az is, hogy a **szórások abszolút értékeinek nagyságrendje számít, amit pedig a változó terjedelme határoz meg**. A hányados tehát annál alacsonyabb lesz, minél nagyobb a vizsgált változó terjedelme.

Azt is hangsúlyozza Hidioglou cikke, hogy az eloszlás pozitív ferdeségének figyelmen kívül hagyása, illetve fel nem ismerése a sokasági paraméterek szisztematikus felülbecslését hozhatja magával, és ezért is körültekintően kell eljárni a megfelelő mintavételi módszer kiválasztása során.

Arra vonatkozóan azonban nincsen támpont a cikkben, hogy mit ért a szerző az extrém aszimmetria fogalmán. Még az sem derül ki, hogy milyen aszimmetria mérőszámot használ fel az aszimmetria mérésére. Ugyanakkor ez lényeges elem, hiszen előrevetíthető, hogy az aszimmetria mértékétől nem független ennek a módszernek a relatív hatásossága. Az aszimmetria mértékének hatására a későbbiekben még visszatérek a számszerű eredmények kapcsán.

Közelítő szabályt találunk az optimális küszöbérték meghatározására, pontosabban egy felső korlátot:

$$y^* = \bar{y} + \sqrt{CV^2 \frac{Y^2}{N} + s^2}$$

ahol:

$y^*$  - az optimális réteghatár felső korlátja (approximate cut-off point)

CV – előírt (elvárt) relatív hibahatár

Ez jó támpontot adhat előzetes számításokhoz, tájékozódáshoz, azonban nem alkalmas a pontos meghatározáshoz. Mindhárom faj esetén sorra vettem néhány önkényesen kiválasztott küszöbértékhez (réteghatárhoz) tartozó eredményt, ahhoz, hogy látni lehessen, nagyságrendileg hol van az optimum, és meghatározható legyen az így szükséges mintavételi arány. Ennél a módszernél tehát a „nagyobb” gazdaságoknál a kiválasztási arány fixen 100 százalék, míg a „kisebbek” esetében az egyszerű véletlen kiválasztást véve alapul a fentiekben meghatározott becslési hiba eléréséhez szükséges elemszám adja meg az eredményt. Ez annyit jelent, hogy minden réteghatár esetén csak 1 változtatható tényező marad: a „kisebb” gazdaságokat tartalmazó rétegen belüli kiválasztási arány. Ennek optimalizálásával jutunk el a szükséges mintanagysághoz. (36. táblázat)

A szarvasmarhákánál a TATS módszerhez tartozó OPS megegyezik az arányosnál talált 15-ös állatállomány nagysággal. A másik két állatfaj esetében jóval alacsonyabb értékek mellett találjuk az optimális réteghatárokat. Sertéseknél 40-50, juhoknál 10-15 állat jelenti az optimumot, attól függően, hogy milyen sokasági szemléletben nézzük. Mindhárom állatfajnál a szűkebb alapsokaság adja a kedvezőbb mintanagyságot, de a szükséges mintavételi arány itt mégis magasabb.

36. TÁBLÁZAT: SZÜKSÉGES MINTANAGYSÁG KÜLÖNBÖZŐ RÉTEGHATÁROK MELLETT (TAKE ALL – TAKE SOME)

Állatfaj	Réteghatár (állatállomány nagyság)	Teljes sokaság		Állattartók	
		n	n/N (%)	n	n/N (%)
Szarvasmarha	50	2 175	1,96	1 517	1,90
	40	1 820	1,64	1 269	1,59
	30	1 431	1,29	1 030	1,29
	20	1 065	0,96	838	1,05
	15	976	0,88	822	1,03
	10	987	0,89	894	1,12
	8	1 087	0,98	1 022	1,28
Sertés	70	1 531	1,38	1 126	1,41
	60	1 442	1,3	1 086	1,36
	50	1 387	1,25	1 070	1,34
	40	1 365	1,23	1 094	1,37
	30	1 664	1,5	1 245	1,56
	20	1 964	1,77	1 812	2,27
Juh	50	6 435	5,8	4 591	5,75
	20	2 563	2,31	2 068	2,59
	15	2 186	1,97	1 852	2,32
	10	2 019	1,82	1 860	2,33
	8	2 053	1,85	1 900	2,38

Forrás: KSH adatbázis alapján saját számítás

Talán furcsának tűnik, hogy a korábbiakban ezeknél alacsonyabb réteghatárok mellett is elvégeztem a számításokat. Ebben az esetben ez azért nem így történt, mert a TATS módszer relatíve alacsony réteghatár mellett már biztosan nem ad optimumot, mert addigra már akkorára duzzad a 100 százalékban megfigyelendő („take all”) egyéni gazdaságok köre, hogy ez már önmagában, a másik réteg nélkül is „túl nagy” mintaelemszámot generál. Ezért tehát túl alacsony értékek mellett már nem is érdemes tovább számolni.

37. TÁBLÁZAT: SEFF KÜLÖNBÖZŐ RÉTEGHATÁROK MELLETT (TATS/EV, TATS/AR)

Állatfaj	Réteghatár (állatállomány nagyság)	SEFF(TATS/EV)		SEFF(TATS/AR)	
		Teljes sokaság	Állattartók	Teljes sokaság	Állattartók
Szarvasmarha	50	0,594	0,837	1,816	1,793
	40	0,497	0,700	1,519	1,500
	30	0,391	0,568	1,194	1,217
	20	0,291	0,462	0,889	0,991
	15	0,267	0,454	0,815	0,972
	10	0,270	0,493	0,824	1,057
	8	0,297	0,564	0,907	1,208
Sertés	70	0,345	0,507	0,677	0,779
	60	0,325	0,489	0,637	0,752
	50	0,313	0,482	0,613	0,740
	40	0,308	0,493	0,603	0,757
	30	0,375	0,561	0,735	0,862
	20	0,443	0,816	0,868	1,254
Juh	50	0,076	0,105	0,124	0,123
	20	0,030	0,047	0,049	0,055
	15	0,026	0,042	0,042	0,050
	10	0,024	0,043	0,039	0,050
	8	0,024	0,044	0,040	0,051

Forrás: KSH adatbázis alapján saját számítás

A mérethatásossági mutatóval érzékelhetjük még jobban, hogy a „take all-take some” módszere mekkora „hasznot” hajthat az egyszerű véletlen mintavétel, illetve az arányos elosztáshoz képest (lásd 37. táblázat). Nézzük meg külön-külön a három állatfaj szerinti eredményeket! A szarvasmarhák hatékony megfigyeléséhez szükséges mintanagyság negyedére-felére csökkent le az EV kiválasztáshoz képest. Viszont az arányos és a TATS elosztás között nem találunk már ennyire markáns különbséget. 10 százalék körüli a minta csökkenésének mértéke. A sertéseknél felére-harmadára csökken a szükséges mintanagyság az EV-hez, és 60-75 százaléka az arányos kiválasztáshoz képest. A sertéseknél tehát nagyobb a „nyereség” a talán egy kicsit bonyolultabb elosztási eljárással. A juhoknál található értékek első ránézésre akár elírásnak is tűnhetnek. Ahogy azonban a Neyman-féle optimális elosztás esetében, itt is az okozza a meglehetősen alacsony értékeket, hogy kiindulásképpen az EV technika mellett a juhoknál 60-75 százalékos szükséges kiválasztási arányt láttunk, de még az arányos rétegzés is közel 50 százalékos mintavételi arány mellett hozta csak a kívánt hibaszintet. Ezek miatt a rendkívül magas viszonyítási értékek miatt látjuk azt, hogy a tizedénél is kevesebbre csökken a szükséges minta nagysága.

Úgy vélem, a Neyman-féle optimális elosztás és a TATS módszer összehasonlítását érdemes külön elvégezni. Hidioglou cikke alapján ugyanis az az előfeltevés fogalmazódott

meg bennem (H5), hogy mivel közepes, illetve erős pozitív aszimmetriát mutat az egyéni gazdaságok sokasága mindhárom állatfajnál, ezért a TATS módszernek hatékonyabbnak kellene lennie a Neyman-féle optimális allokációnál.

38. TÁBLÁZAT: SEFF KÜLÖNBÖZŐ RÉTEGHATÁROK MELLETT (NO/TATS)

Állatfaj	Réteghatár (állatállomány nagyság)	SEFF(NO/TATS)	
		Teljes sokaság	Állattartók
Szarvasmarha	30	0,471	0,475
	20	0,899	0,860
	15	0,761	0,667
	10	0,602	0,515
Sertés	20	0,393	0,312
Juh	10	-	0,978
	8	0,973	0,975

*Forrás: KSH adatbázis alapján saját számítás*

A mérethatásossági mutató akkor támasztja alá a feltevésemet, ha 1-nél nagyobb értékeket vesz fel. A 38. táblázatra tekintve azonban nem ezt tapasztaljuk. A juhok kivételével egynek a közelében sincsenek az értékek. A 39. táblázatból láthatjuk (ahogy a leíró elemzéseknél is megállapítottuk), hogy a szarvasmarhák és juhok esetében szinte egyforma mértékű aszimmetriával találkoztunk. Ez azt jelenti, hogy az aszimmetria mértéke önmagában nem tehető felelőssé a különbségekért, vagyis csak az erős aszimmetria nem oka annak, hogy a TATS hatékonyabb a Neyman-féle optimális elosztásnál. Ugyanezt látjuk akkor is, ha nem réteghatárok mentén végezzük el az összehasonlítást, hanem mindkét technika mellett kiválasztjuk a szükséges mintanagyság minimumát és ezt viszonyítjuk egymáshoz a SEFF mutatón keresztül (39. táblázat). Világosan látszik, hogy a szarvasmarhánál közel 90 százalékkal, a sertésnél pedig több mint 50 százalékkal bizonyul hatékonyabbnak a Neyman-féle optimális elosztás. Juhoknál gyakorlatilag egyforma eredményeket érhetünk el.

39. TÁBLÁZAT: LEGALACSONYABB SZÜKSÉGES MINTAELEMSZÁMOK ÖSSZEHASONLÍTÁSA

Állatfaj	SEFF(NO/TATS)		Ferdesség ( $S_2$ mutató)	
	Teljes sokaság	Állattartók	Teljes sokaság	Állattartók
Szarvasmarha	0,114	0,118	0,953	0,958
Sertés	0,415	0,433	0,708	0,645
Juh	0,989	0,983	0,950	0,942

*Forrás: KSH adatbázis alapján saját számítás*

Tehát az erőteljes jobb oldali aszimmetria önmagában nem magyarázza a TATS mintavételi módszer hatékonyságát. Ahogy a bevezetőben, az **5. hipotézisben** (H5) megfogalmaztam, úgy véltem, hogy az aszimmetria mellett a szórásnak is fontos szerepe van. Egy dolog van a ferdeség mellett, amiben a szarvasmarhák és juhok eloszlása jelentősen eltér egymástól, és ez nem más, mint a terjedelem, illetve ennek okaként a szórás eltérő abszolút nagysága. A szarvasmarháknál a terjedelem nem érte el a 100-at, így a juhoknál közel 1000 ez az érték. Ebből, és az erős aszimmetriából együttesen következik, hogy a szórások értéke is nagyon magas. És ami ennél is fontosabb, hogy az erős aszimmetria miatt hiába „vágjuk” állatállomány méretük szerint a két részre a sokaságot, a szórás a nagyságkategóriákban is viszonylag magas marad, ami nagyban meghatározza a mintavételi módszerekkel elérhető hatékonyságot. A H5 az empirikus adatok alapján igazolást nyert, és a következő tézis fogalmazható meg:

## 6. tézis

**A „take all – take some” mintavételi technika és a Neyman-féle optimális elosztás (2 réteget feltételezve) egymáshoz viszonyított hatékonyságát a vizsgált változó aszimmetriája és terjedelme határozza meg.**

### 5.2.4. *További lehetőségek az OPS meghatározásához*

Az optimális réteghatár(ok) meghatározásának elméleti szabályszerűségeit többen próbálták már leírni, különböző kiinduló feltevések mellett, különböző hatékonysággal. (Fraller 2011) Az állatállomány példáján további két módszert szeretnék bemutatni, illetve kipróbálni. Mindkettő lényege, hogy olyan osztópontokat, réteghatárokat igyekszik meghatározni, amelyek a variancia minimalizálására törekszenek. Az egyik a szélesebb körben elterjedt Dalenius-Hodges szabály, a kumulatív  $\sqrt{f}$  szabály (Dalenius – Hodges 1959), a másik pedig a kissé bonyolultabb Ekman-szabály (Ekman 1959). Többek között Cochran (1961) is végzett számításokat, amelyek összehasonlítják a kétféle módszerrel kapott eredményeket. Arra a következtetésre jutott, hogy az Ekman-szabály adta a legjobb megoldást aszimmetrikus eloszlással jellemezhető sokaságokra. Az Ekman-szabály az elérhető leghatékonyabb rétegzéshez képest is jól szerepelt, vagyis a gyakorlatban kimutatható



optimumhoz a legközelebbi eredményeket adta. Ennyi felvezetés után bemutatom részletesen a két módszer sajátosságait, és az állatállomány példáján a konkrét számszerűsítés menetét. Mivel az eddigi számításaimban két rétegben gondolkodtam, és erre vannak optimális réteghatár eredményeim, mindkét módszert két rétegre alkalmazom. A korábbiakban bemutatott Neyman-féle optimális allokációhoz tartozó OPS-ekkel összehasonlítva értékelni is lehet majd a kapott értékeket.

### **Dalenius-Hodges (kumulatív $\sqrt{f}$ ) szabály**

Ahhoz, hogy a módszert alkalmazni tudjuk az alapsokaságot egyenlő hosszúságú osztályközös gyakorisági sorba kell rendezni, ahol az osztályközök száma viszonylag nagy (legalább 15-20). Az osztályközök melletti gyakoriságok ( $f$ ) adják a számítások alapját. Minden osztályköz mellé kiszámítjuk a hozzá tartozó gyakoriság négyzetgyökét, és ezeket az értékeket kumuláljuk felülről lefelé. A módszer feltételezi, hogy eldöntöttük előre, hány réteget szeretnénk. A következő egyenlőséget kell minél pontosabban teljesíteni, közelíteni:

$$\sum_{j=1}^{J_1} \sqrt{f_j} = \sum_{j=J_1+1}^{J_2} \sqrt{f_j} = \dots = \sum_{j=J_{H-2}+1}^{J_{H-1}} \sqrt{f_j} = \sum_{j=J_{H-1}+1}^J \sqrt{f_j}$$

ahol:

$J$  – az osztályközök száma

$j=1, 2, \dots, J$

$f_j$  – a  $j$ -dik osztályközhez tartozó gyakoriság

$H$  – a létrehozni kívánt rétegek száma

A fenti formula azt jelenti, hogy a gyakoriságok négyzetgyökének összegét annyi részre osztjuk, ahány réteg kialakítását tervezzük. Az így kapott értéket (illetve ennek többszörösét) keressük meg a kumulált  $\sqrt{f}$  oszlopban. Ahová ezek az értékek esnek (legalább közelítőleg), ott kell meghúznunk a réteghatárokat. Vagyis a megfelelő  $J_H$  értékeket, esetünkben csak  $J_1$ -et kell ilyen módon megtalálni.

A 40-42. táblázatok tartalmazzák a három vizsgált állatfaj gyakorisági tábláit és a kumulatív  $\sqrt{f}$  szabály alkalmazásához szükséges értékeket. (A gyakorisági soroknak csak töredékét tüntettem fel az áttekinthetőség kedvéért. A teljes táblákat a VIII. számú melléklet tartalmazza.)

40. TÁBLÁZAT: SZARVASMARHÁT TARTÓ GAZDASÁGOK GYAKORISÁGI SORA

Osztályköz (db)	Gyakoriság (f)	$\sqrt{f}$	Kumulált $\sqrt{f}$
0-5	78 483	280,1	280,1
6-10	656	25,6	305,8
11-15	222	14,9	320,7
16-20	136	11,7	332,3
21-25	93	9,6	342,0
...	...	...	...
86-90	1	1,0	391,4
91-95	2	1,4	392,8

*Forrás: KSH adatbázis alapján saját számítás*

41. TÁBLÁZAT: SERTÉST TARTÓ GAZDASÁGOK GYAKORISÁGI SORA

Osztályköz (db)	Gyakoriság (f)	$\sqrt{f}$	Kumulált $\sqrt{f}$
0-10	75721	275,2	275,2
11-20	2599	51,0	326,2
21-30	737	27,1	353,3
31-40	294	17,1	370,4
41-50	135	11,6	382,1
...	...	...	...
461-470	0	0,0	465,1
471-480	1	1,0	466,1

*Forrás: KSH adatbázis alapján saját számítás*

42. TÁBLÁZAT: JUHOT TARTÓ GAZDASÁGOK GYAKORISÁGI SORA

Osztályköz (db)	Gyakoriság (f)	$\sqrt{f}$	Kumulált $\sqrt{f}$
0-10	78335	279,9	279,9
11-20	490	22,1	302,0
21-30	189	13,7	315,8
31-40	125	11,2	326,9
41-50	97	9,8	336,8
...	...	...	...
941-950	0	0,0	502,0
951-960	2	1,4	503,4

*Forrás: KSH adatbázis alapján saját számítás*

Mivel két rétegben gondolkodunk, mindhárom esetben a táblázat jobb alsó sarkában lévő  $\sqrt{f}$  összeget kell két részre osztani és megnézni a kumulált  $\sqrt{f}$  értékei között, hogy melyikhez esik a legközelebb a kapott szám. Mindhárom állatfajnál az első sorban találjuk az optimumot. Vagyis a használt osztályközök szerint szarvasmarhánál 5, sertésnél és juhnál pedig a 10 lesz az OPS. Ha visszatekintünk a korábban elért eredményekhez, akkor láthatjuk, hogy a Neyman-féle optimális elosztás mellett hasonló eredményekre jutottunk (csak jóval több számítás és iteráció után).

Mindemellett a módszer legnagyobb problémája is megvilágítást nyer a számpéldákon keresztül. Nevezetesen az, hogy nincsen elmélet arra vonatkozólag, mi a J optimális, legjobb értéke, vagyis hogy kiindulásnak hány osztályközbe rendezzük a sokaságot, milyen hosszúságú osztályközöket alkalmazzunk. Erre áthidaló megoldás lehet, ha a lehető legkisebb osztályközöket alakítjuk ki. Példánkban, az állatállománynál ez kettes hosszúságú intervallumokat jelentene. Ez a mai technikai feltételek mellett nem jelent többlet erőforrás-ráfordítást, hiszen a szoftvernek „mindegy”, hogy mekkora osztályközökbe helyezi el a sokaságot. Természetesen a vizsgált paraméter nagyságrendjétől függ, hogy mi az észszerűség határán belüli legkisebb intervallumhossz. Árbevételre vonatkozó összeírásoknál nyilván nem lehet 2 forintos osztályközöket kialakítani, hiszen nem is veszik fel ilyen részletességgel az adatokat. Ott a 2-5 millió forintos intervallumhossz jelentheti a legkisebb szóba jöhető értéket. A legsűrűbb osztályközök alkalmazása viszont elfogadható pontosságú értéket ad az OPS-re, és ezzel a kumulatív  $\sqrt{f}$  szabály gyakorlati előnyei vitathatatlanok.

### **Ekman-szabály**

Az Ekman-szabály azt mondja ki, hogy a réteghatároló pontokat aszerint válasszuk ki, hogy minél jobban megfeleljenek a következő összefüggésnek:

$$N_1(b_1 - b_0) = N_2(b_2 - b_1) = \dots = N_H(b_H - b_{H-1})$$

ahol:

$h = 1, 2, \dots, H$  a kialakítani kívánt rétegek száma

$N_h$  – a H-dik réteg alapsokasági elemszáma

$b_h$  – a H-dik réteget határoló felső korlát értéke ( $b_0$  = sokasági értékek minimuma,  $b_H$  = a sokasági értékek maximuma)

Csak a legjobb közelítést találhatjuk meg az egyenlőségek kielégítésére, hiszen pontos megoldás a gyakorlatban nem feltétlenül létezik, amennyiben a rétegek elemszáma egész szám. Márpedig az összeírások során egész egységekkel dolgozunk. (Létezik az Ekman-szabálynak egy kiterjesztése is, ami nem-egész rétegnagyságok esetére mutatja be a pontos megoldást viszonylag általános feltételek mellett. [Hedlin 2000] Mivel azonban a dolgozat, és a gyakorlati összeírások szempontjából ez a helyzet nem áll fenn, ezt nem mutatom be.)

43. TÁBLÁZAT: EKMAN-SZABÁLY SZÁMÍTÁSA A SZARVASMARHÁKNÁL

Rétegek (osztályközök)	Alapsokasági elemszám ( $N_h$ )	$N_h(b_h - b_{h-1})$
0-5	78 483	392 415
6-93	1 342	116 754
0-3	77 924	233 772
4-93	1 901	169 189
0-2	77 507	155 014
3-93	2 318	208 620

*Forrás: KSH adatbázis alapján saját számítás*

44. TÁBLÁZAT: EKMAN-SZABÁLY SZÁMÍTÁSA A SERTÉSEKNÉL

Rétegek (osztályközök)	Alapsokasági elemszám ( $N_h$ )	$N_h(b_h - b_{h-1})$
0-10	75 721	757 210
11-473	4 104	1 896 048
0-14	77 209	1 080 926
15-473	2 616	1 198 128
0-15	77 519	1 162 785
16-473	2 306	1 069 984

*Forrás: KSH adatbázis alapján saját számítás*

45. TÁBLÁZAT: EKMAN-SZABÁLY SZÁMÍTÁSA A JUHOKNÁL

Rétegek (osztályközök)	Alapsokasági elemszám ( $N_h$ )	$N_h(b_h - b_{h-1})$
0-10	78 335	783 350
11-955	1 490	1 406 560
0-12	78 482	941 784
13-955	1 343	1 265 106
0-14	78 568	1 099 952
15-955	1 257	1 181 580
0-15	78 630	1 179 450
16-955	1 195	1 122 105

*Forrás: KSH adatbázis alapján saját számítás*

Ránézésre ez a módszer sem kíván túl sok számítást, hiszen egyszerű sokasági adatokra van mindössze szükségünk. Azonban, ha belegondolunk, akkor ez egy komoly iterációs probléma, mert folyamatosan mozgatni kell a határoló pontokat, amíg az egyenlőséget elégségesen közelíteni nem tudjuk. Ráadásul az „elégséges közelítés” kritériumát sem tudjuk pontosabban definiálni, ezért biztosan több próbálkozásra van szükség, hogy dönteni lehessen.

A 43. táblázat-45. táblázatok szemléltetik az Ekman-szabály alkalmazásának menetét. Két réteget megcélözva gyakorlatilag 3-4 lépésben el is jutottunk a lehető legjobb OPS meghatározásához. A fenti egyenlőség szinte természetes módon sehol nem teljesül, azt ténylegesen csak közelíteni lehetett. A sertés és juh állatfajoknál 5-8 százalékos eltérés maradt a szorzatok között, a szarvasmarháknál viszont 25 százalékos rést voltunk kénytelenek „optimumnak” elfogadni. Ez a három konkrét példa azt illusztrálja, hogy a „lehető legjobb közelítés” mértéke nagyon is eltérő lehet. Ez két dologra hívja fel a figyelmet. Egyrészt nem szabad elvárni, hogy az egyenlőség teljesüljön, bizonyos esetekben még közelítőleg sem. Másrészt azt a következtetést is levonhatjuk, hogy a próbálkozás kezdetén, ha az egyenlőségben szereplő tagok között viszonylag nagynak érzékelt távolság van, akkor sem szabad megijedni, hiszen az „optimum” akár pár lépésre is lehet. Az iterációs lépések száma ugyanakkor nagymértékben függ attól, hogy hány réteget szeretnénk kialakítani. Két réteg esetén csak egy réteghatárt mozgatunk. Ahogy növeljük a rétegek számát, úgy nő a mozgatandó réteghatárok száma is, és úgy bonyolódik a helyzet.

Ami a konkrét eredményeket illeti, azt kell megállapítani, hogy az Ekman-szabály nem adta vissza minden esetben a korábbiakban elért optimális értékeket. Szarvasmarhára szépen

kijött a 2 állatot mutató réteghatár. A sertéseknél 9-10 állatnál volt a határ és nem 15-nél, juhoknál pedig hasonlóan 10 állatnál találtuk meg korábban az optimumot és nem 15-nél.

A magyarországi állatállomány példáján, a konkrét számításokon keresztül tehát azt látjuk, hogy a Dalenius-Hodges féle kumulatív  $\sqrt{f}$  szabály adta vissza pontosabban a tényleges optimumot, míg a Ekman-szabály csak egy esetben volt pontos, a másik két esetben pontatlanabb volt.

Mindkét módszer vitathatatlan előnye, hogy nem igényel túlságosan bonyolult számításokat. Persze szükséges a sokaság egyedi értékeinek ismerete az osztályközök és elemszámok kiszámításához. Ezek meghatározása azonban mégis egyszerűbb, mint a szórások számszerűsítése. A mintaelemszámot is mindkét módszer kidolgozása során előre adottnak tételezték fel a szerzők, és emellé kerestek minimális varianciát adó réteghatár(oka)t. A módszerek gyakorlati alkalmazásához ugyanakkor nincsen szükség rá, hogy előre meghatározzuk a mintavételi arányt. A fent bemutatott számításokhoz nem használtam fel a mintanagyságot, mégis eljutottam az optimális megoldáshoz (legalábbis annak egyfajta közelítéséhez). A módszerek további előnye, hogy tetszőleges számú réteg optimális kialakítására adnak lehetőséget, így tulajdonképpen a rétegek számát sem feltétlenül kell előre eldönteni, hanem többféle rétegszám mellett is relatíve könnyedén meghatározhatóak a határoló pontok, és elvégezhetőek a további számítások a mintavételi terv kialakításával kapcsolatban. Az állatállomány adatbázisán végzett kísérleti számítások alapján mégis azt kell mondanom, hogy a kumulatív  $\sqrt{f}$  vizsgázott jobban. Egyszerűbb az alkalmazása, különösen akkor, ha kettőnél több réteget szeretnénk kialakítani, illetve ha többféle rétegszámot szeretnénk kipróbálni. Bármilyen is a célunk, mindvégig ugyanabból az – lehető legrészletesebb – osztályközös gyakorisági sorból indulunk ki, és nincsen szükség számottevő további kalkuláció elvégzésére. Az Ekman-szabály nehezkesebb, rugalmatlanabb, és a bemutatott példákat tekintve kevésbé hatékony módszernek bizonyult.

### **OPS meghatározás exponenciális eloszlással jellemezhető sokaságra**

Hunyadi (1991) cikkében szintén említést tesz Dalenius és Hodges, valamint Ekman által kidolgozott módszerekről. Azt is hozzáteszi, hogy a gyakorlatban igazából sosem áll rendelkezésre naprakész információ magáról a vizsgált változóról egyedenként, ezért önmagában mindkét módszer csak elméleti szinten valósítható meg. Ugyanakkor minden

körülmények között lehetőségünk nyílt beszerezni a szükséges információt egy előzetes, egyszerű, kisebb volumenű (elő)összeírás lebonyolításával. Ezt nevezzük kétfázisú mintavételnek. Ilyenkor az első fázis célja pusztán a rétegzéshez szükséges „külső” információ megszerzése.

A tanulmányban Hunyadi egy speciális esetet feltételezve mutatja be az optimális réteghatár meghatározásának menetét, és az elérhető hatékonyságjavulás mértékét. Abból indul ki, hogy a vizsgált változó exponenciális eloszlást követ. Jelen dolgozatnak nem célja a cikk részletes ismertetése, csak a főbb eredményeket szeretném bemutatni, és a módszerben rejlő lehetőségeket.

Bizonyítja, hogy az OPS meghatározásához mindössze az exponenciális eloszlás paraméterét kell ismerni, nagysága mástól nem függ. Ezek alapján exponenciális esetben az optimális réteghatár (2 réteget feltételezve) a következő összefüggéssel nyerhető:

$$OPS = 1,594 \frac{1}{\lambda}$$

Azt is bemutatja a cikk, hogy az így kiválasztott mintából történő becslés relatív hatásossága (DEFF) az egyszerű véletlenhez képest 0,2855, amennyiben teljesen pontosan ismerjük az exponenciális eloszlás paraméterét. Mivel azonban a paraméter előzetes ismerete nem reális feltételezés, ezért arról is képet kapunk, hogy milyen módon határozhatjuk meg a kétfázisú mintavételhez szükséges mintaelemszámokat meghatározott költségkeret mellett. Táblázatba foglalva jelenik meg a DEFF mutató értéke különböző mintanagyságok mellett. 100 ezres minta esetén már szinte elérhető az elméleti hatékonyság javulás mértéke.

Ahogy korábban már magam is kitértem erre, nem tűnik elképzelhetetlennek, hogy a gazdaságstatisztikában előfordulnak ilyen eloszlású változók. Éppen ezért meg is vizsgáltam a három állatfaj gyakorisági sorának segítségével, hogy megfelelően közelíthetőek-e exponenciális eloszlással. A sertés és a juh esetében nem jártam sikerrel, viszont a szarvasmarha állomány (a szarvasmarhát nem tartó gazdaságok nélkül) szinte tökéletesen illeszkedik az exponenciális eloszláshoz a kísérleti adatbázis adatai alapján. Az exponenciális eloszlás paraméterét nagyon egyszerűen megkaphatjuk, ha a sokasági várható érték reciprokát vesszük. Ennek megfelelően  $\lambda=0,123$  eredmény alakult ki. Kiszámítottam az elméleti eloszlás sűrűségfüggvényének értékeit a kapott paraméterrel. Az eredeti gyakorisági sor relatív gyakoriságaival kellett összevetni az eredményeket. Az illeszkedésvizsgálathoz Khi-

négyzet próbát alkalmaztam. Az értékek alapján (lásd 46. táblázat) minden ésszerű szignifikancia szint mellett el kellett fogadni, hogy a szarvasmarha állomány exponenciális eloszlást követ. A számításhoz felhasznált teljes gyakorisági sor a X. számú melléklet tartalmazza.

46. TÁBLÁZAT: AZ EXPONENCIÁLIS ILLESZTÉS FŐBB EREDMÉNYEI

Mutató	Érték
Teljes értékösszeg	28 032
Elemszám	3 447
Átlag (várható érték)	8,132
lambda	0,123
Számított Khi-négyzet érték	0,819
Kritikus Khi-négyzet érték	98,484
p-érték	1,000

*Forrás: Saját számítás*

$$OPS = 1,594 \frac{1}{0,123} = 12,94$$

Ez azt jelenti, hogy a 13 alatti állatállománnyal rendelkező gazdaságok alkotják a „kicsik” rétegét, és a 13 vagy afeletti állomány a „nagyok” közé sorolja a gazdákat. Az arányos elosztáshoz korábban a 15-ös réteghatár bizonyult optimálisnak a vizsgálataim során. Viszont én nem végeztem el a számítást minden egyes pontban, így Hunyadi megoldásával pontosítottuk ezt az értéket. És ezzel egyúttal azt is beláthatjuk, hogy ez valóban az optimumot hozta ki mindössze két lépésben.

Ezek alapján kimondhatjuk, hogy nem haszontalan extrém ferde eloszlásoknál élni azzal a gyanúval, hogy esetleg exponenciális eloszlást követnek, vagy legalábbis azzal jól közelíthetőek. Maga a paraméter meghatározás és az illeszkedésvizsgálat nem igényel túl sok számítást. Hogyha viszont igazolni tudjuk az illeszkedést, akkor jelentősen megkönnyítjük az OPS meghatározás menetét. Bár a sertések és a juhok esetében az exponenciális eloszlás nem illeszkedik megfelelő, meg lehetne próbálni más nevezetes eloszlásokat, illetve a sokaság „darabolása” is megoldás lehetne. A későbbi kutatások során ezzel is szeretnék még foglalkozni.



### 5.2.5. Rétegzés régiók és nagyságkategóriák szerint

A végső célunk az, hogy az előírásoknak megfelelően régiós szintre reprezentatív, ugyanakkor a pontossági követelményeknek eleget tenni képes hatékony, legkisebb szükséges mintanagysággal megvalósítható mintavételi tervet állítsunk össze. Ennek eléréséhez kombinálni kell a régiós (technikai) és a nagyságkategóriák szerinti (kapcsolati) rétegzést.

A számításokat már csak az egyéni gazdaságok teljes sokaságát alapul véve végzem el, hiszen a korábbiakban beláttuk, hogy az alapsokaság állattartó gazdaságokra való szűkítésének nincsen elegendő hozadéka.

A szimulációk számításigénye ezzel ugrásszerűen megemelkedett. Minden egyes réteghatár mellett 14 rétegre kell kiszámítani az elemszámokat, a szórásokat. Ráadásul a szórás számítása, ahogy már említettem, meglehetősen bonyolult, hiszen az adatbázisban csak azok a gazdaságok szerepelnek, akik rendelkeznek valamilyen állatállománnyal. Ezért az állatállománnyal, illetve adott fajjal nem rendelkező gazdaságok számát külön kell figyelembe venni, és az átlagot, illetve a szórást ezekkel együtt kell átszámítani.

A szemléltetés, és az egyszerűség kedvéért a korábban bemutatott optimális réteghatárokon mentén kezdtem el a számításokat. Ezek mellett csak néhány réteghatárt vizsgáltam meg.

A rétegek száma így  $2 \times 7$ -re vagyis 14-re növekszik a régiók és a nagyságkategóriák kombinálásával. Ezúttal is állatfajonként néztem végig az eredményeket, és határoztam meg a szükséges legkisebb mintavételi arányt. Mivel korábban már beláttuk, hogy a régiós (technikai) rétegzés jelentősen nem javít a mintavétel hatékonyságát, már a számítások előtt világos volt, hogy a nagyságkategóriák alkalmazásánál kialakult nagyon alacsony mintavételi arányok növekedni fognak a két rétegzési szempont egyidejű alkalmazásával.

Éltem azzal előfeltevéssel, hogy a Neyman-féle optimális elosztás hatékonyabban működik erre a sokaságra, ezért itt elsősorban erre koncentráltam. A konkrét számítások elvégzése során azonban minden egyes esetben azt kellett tapasztalnom, hogy a „nagyobb” gazdaságokat tartalmazó rétegeknél az alapsokasági elemszámot ( $N_h$ ) meghaladó mintanagyságot ( $n_h$ ) igényelt volna a szórások alapján optimalizáló allokáció. Vagyis, gyakorlatilag egy „take all – take some” típusú mintavétel valósult meg. Mind a 7 régióban a

„nagyobb” gazdaságok kiválasztási aránya  $(\frac{\sum_{h=1}^7 n_h}{\sum_{h=1}^7 N_h})$  100 százalék lett, és csak a „kisebbeket”

tartalmazó 7 rétegben zajlik valódi mintavétel. Ez a helyzet pedig új fénybe helyezte az eddigi elgondolásokat. Így már érdemes volt megvizsgálni az arányos rétegzés lehetőségét is, hiszen már csak 7 rétegben használtam a tényeleget optimalizálást, és a rétegenkénti kiválasztási arányok sok esetben nem mutattak túl nagy eltérést (szóródást). Ráadásul az arányos rétegzés sokkal egyszerűbb is, és nem igényli a szórások használatát így ismeretét sem.

Az eredmények szemléltetéséhez egy olyan táblázatot szerkesztettem (47. táblázat), amiben összevontan szerepel a réteghatár feletti (1-7. rétegek) és alatti gazdaságok (8-14. rétegek) száma (régiók szerinti bontás nem látszik). Ezzel külön is érzékeltetni szeretném, hogy melyik rétegekből milyen a számosság. Természetesen a mintavételi arányok is érdekesek. Külön szerepel Neyman-féle optimális elosztás mellett a „kicsik” rétegeinek átlagos kiválasztási aránya, a teljes sokaságra vonatkozó kiválasztási arány. A utolsó két oszlop pedig az arányos elosztás mellett érvényesülő kiválasztási arányokat tartalmazza. (A „nagyobbak” rétegeiből itt is teljes körű kiválasztás van.)

47. TÁBLÁZAT: SZÜKSÉGES MINTANAGYSÁG RÉGIÓS ÉS NAGYSÁGKATEGÓRIÁS RÉTEGZÉS MELLETT

Állatfaj	Réteghatár	„Nagyok” száma <sup>39</sup>	Neyman-féle optimális		Arányos	
			A „kisebbek” száma (kiválasztási arány) <sup>40</sup>	n/N	A „kisebbek” száma (kiválasztási arány)	n/N
Szarvasmarha	2	2 318	14 (0,01%)	2,07%	95 (0,09%)	2,14%
	20	328	4 560 (4,17%)	4,45%	4 306 (3,90%)	4,19%
	50	69	12 433 (11,24%)	11,29%	11 731 (10,60%)	10,66%
Sertés	6	6 998	70 (0,07%)	6,38%	374 (0,36%)	6,66%
	10	4 004	297 (0,28%)	3,88%	854 (0,80%)	4,39%
	20	1 505	1 692 (1,55%)	2,89%	2 622 (2,40%)	3,73%
	30	768	3 094 (2,81%)	3,49%	3 849 (3,50%)	4,17%
	100	94	9 257 (8,34%)	8,44%	9 073 (8,20%)	8,28%
Juh	8	1 735	91 (0,08%)	1,65%	1 853 (3,24%)	3,24%
	50	589	14 891 (13,52%)	13,98%	21 131 (21,00%)	21,42%

Forrás: KSH adatbázis alapján saját számítás

Az eredmények, úgy vélem, még mindig ígéretesek. Állatfajonként a legkisebb kiválasztási arányt a színezett cellák jelölik. Szarvasmarha esetén azért nem a Neyman-féle optimális allokáció melletti 2,07 százalék színes, mert itt a 8-14. rétegekből összesen 14 elem kerülne a mintába, ami nem biztosítja a minimálisan szükséges számosságot az összes rétegből, ezért a gyakorlatban nem kivitelezhető.

Számomra mindenképpen meglepő, hogy a legalacsonyabb érték a juhoknál szerepel. Ez az eredmény támasztja alá leginkább azt a tényt, hogy nagyon megéri a nagyságkategóriák szerinti rétegzéssel kísérletezni, mert elképesztően hatékony mintavételi tervet lehet így kialakítani.

$$\sum_{h=1}^7 n_h$$

$$\sum_{h=8}^{14} n_h \quad \text{és zárójelben: } \left( \frac{\sum_{h=8}^{14} n_h}{\sum_{h=8}^{14} N_h} \right)$$

A dolgozatban mindvégig külön-külön vizsgáltam meg az egyes állatfajok jellemzőit, és külön határoztam meg a mintavételi optimumokat. Ezzel tehát önmagában még nem adtam meg a teljes megoldást, hiszen az összeírásoknál végső soron egy mintasokaság van, amit lekérdeznek, és ebből kell minden állatfajra a hatékony becsléseket elvégezni. Ez azt jelenti, hogy a kapott „részeredményeket” valahogy „össze kellene gyúrni” egy egésszé, egy mintavételi tervé. Ehhez azt a megoldást látom a legígéretesebbnek, hogy a mintavétel előtt a gazdaságokat első körben két részre osztjuk: „nagyobbakra” és „kisebbekre”. Azok lesznek a nagyok, akik a fent meghatározott réteghatárok közül legalább az egyiket elérik. A „kisebbekből” akkora egyszerű véletlen mintát választunk ki, amekkora a három állatfaj eredményei szerint a maximum, vagyis 1692-t. A „nagyobbak” csoportját pedig teljes körűen megfigyeljük, így ott további kiválasztásra nincsen szükség. Gyakorlatilag egyfajta „take all – take some” elv kerül alkalmazásra, és a viszonylag nagy egyéni gazdaságokat figyeljük meg 100 százalékos kiválasztási aránnyal. Ez maximális esetben, ha semmilyen átfedés nincsen a „nagyok” csoportjában, és minden gazdaság csak 1 állatfaj „miatt” került be, akkor a korábbi értékek összegéből adódik: 5558 gazdaság. Ezzel tehát egy felső közelítést adjuk a szükséges mintanagyságnak, ami ugyanakkor egészen biztosan alacsonyabb ennél, hiszen korábban láttuk, hogy átfedés nagyon is jellemzőek az állattartásban. A kiválasztási arány így, összességében 5 százalékos. Tisztában vagyok azzal, hogy ezáltal nem kapjuk vissza a fent bemutatott értékeket, hiszen az átlapolások eltéréseket okoznak a szórásokban, és ezáltal a becslési hibában is, de ez már egy jó közelítése az optimális megoldásnak. De ezzel még mindig nem jutottunk el a végleges eredményhez. Ehhez egy további fontos lépés szükséges, amelyet a következő fejezet tartalmaz.

### ***5.3. Módszertani megfontolások a kiválasztási aránnyal kapcsolatban***

A minta elemszámát meghatározza általában a rendelkezésre álló pénzügyi források nagysága és az adatok elvárt pontossága (és ezen felül a felhasznált mintavételi módszer). Ennek fontos (a mintavételes összeírásoknál gyakran önmagában hangsúlyozott) jellemzője a mintavételi arány ( $f=n/N$ ). Ez egy megoszlási viszonzyszám, ami a minta nagyságát mutatja meg százalékos formában az alapsokasági elemszámhoz viszonyítva. Általánosságban elmondható, hogy minél nagyobb ez a mintavételi arány ( $f$ ), annál pontosabban lehet következtetni az alapsokasági jellemzőkre. Ugyanakkor nem szabad megfeledkeznünk arról sem, hogy emellett az alapsokaság és a minta konkrét nagysága, számossága is igen fontos

tényezők. Egy országos 1 százalékos mintából például pontosabb eredményeket kaphatunk, mint egy tízszer akkora mintavételi arány esetén (10 százalékos) egy megyei mintából. Ennek oka mindössze a sokaságok nagyságrendjében keresendő (hazánkban általában feltételezhető, hogy a vizsgált társadalmi jelenségek országosan és területi bontásban is hasonló szóródással jellemezhetőek.) A becslés hibájának számszerűsítésekor ugyanis a minta elemszáma (illetve annak négyzetgyöke) a nevezőben szerepel, és így jelentősen befolyásolja annak nagyságát. Ezért tehát a kiválasztási arányt sosem szabad elszigetelten kezelni, mindig meg kell nézni hozzá a sokaság konkrét számosságát is.

Mindezt azért tartottam fontosnak külön fejezetpontban kifejteni, mert a szimulációs eredmények közvetlenül nem hasonlíthatók össze a KSH-ban aktuálisan alkalmazott mintavételi módszerrel a hatásosság szempontjából. A mérőhatásossági mutató ugyanis csak akkor használható, ha ugyanarról az alapsokaságról beszélünk. A szimulációs számítások alkalmával azonban én egy szűkebben értelmezett alapsokasággal dolgoztam, ami így a közvetlen összehasonlítást nem teszi lehetővé. A **mintavételi arányok** ugyan minden esetben **százalékos értékek, de nem mindegy, hogy mi a viszonyítási alap**. Az a tény ugyanis, hogy egy 3 százalékos mintavételi arány elegendő a sertésállomány megfigyelésére egy 100 ezres alapsokaság esetén, még nem jelenti azt, hogy ez a mintavételi arány ugyanolyan hatásos lesz egy 600 ezres alapsokasági elemszám mellett. Még akkor sem, ha a becsült szórások változatlanosságát feltételezzük.

Ha egyszerűen átszámítjuk a kísérleti adatbázis elemszámait a teljes alapsokaságra (minden elemszámot beszorzunk a  $618611/110494$  hányados értékével), akkor megkapjuk, hogy melyek lesznek a „tényleges” mintavételi arányok (vagy legalábbis azok jó közelítései) az optimális régió-nagyságkategória kombináció mellett. Ezeket az értékeket kell összevonnunk egy egységes mintává, és az ebből kialakuló mintavételi arány lesz az, amit össze tudunk vetni a jelenleg alkalmazott KSH-s minta méretével. Ennek fényében tudjuk megítélni az itt kidolgozott módszer hatékonyságát, tényleges gyakorlati alkalmazhatóságát.

A régiós-nagyságkategóriás rétegzés eredményeit számítottam át a valódi teljes alapsokaságra is. Azért csak ezt, mert a végső eredménye a mintavételi tervek sorának ez lett. A 48. táblázat formátuma és tartalma megegyezik a 47. táblázatával, annyi a különbség, hogy más elemszámokra (a teljes alapsokaságra) készült el.

A két táblázat eredményeit összehasonlítva láthatjuk, hogy a sokaság elemszámának nagyságrendi változása, a szórások és a belső struktúra változása nélkül is nagyon eltérő eredményeket hoz. Ezért is tartom fontosnak az elemszámok nagyságrendjét is az összehasonlítások elvégzése során. Egy részsokaságra (kísérleti állományra) kapott értékek iránymutatóak, módszertani szempontból kijelölik a hatékony irányt, de közvetlenül nem adják meg a szükséges kiválasztási arányokat.

48. TÁBLÁZAT: SZÜKSÉGES MINTANAGYSÁG RÉGIÓS ÉS NAGYSÁGKATEGÓRIÁS RÉTEGZÉS MELLETT A VALÓS ALAPSOKASÁGRA BECSÜLVE (N=618 500)

Állatfaj	Réteghatár	„Nagyok” száma <sup>41</sup>	Neyman-féle optimális		Arányos	
			A „kisebbek” száma (kiválasztási arány) <sup>42</sup>	n/N	A „kisebbek” száma (kiválasztási arány)	n/N
Szarvasmarha	2	12 925	14 (0,002%)	2,09%	94 (0,02%)	2,10%
	20	1 829	4 834 (0,79%)	1,10%	4 433 (0,72%)	1,01%
	50	385	14 107 (2,29%)	2,35%	12 959 (2,10%)	2,16%
Sertés	6	39 021	116 (0,02%)	6,34%	602 (0,10%)	6,42%
	10	22 326	405 (0,07%)	3,68%	1 154 (0,19%)	3,80%
	20	8 392	1 694 (0,28%)	1,63%	2 680 (0,44%)	1,79%
	30	4 282	3 181 (0,52%)	1,21%	4 047 (0,66%)	1,35%
	100	524	10 020 (1,62%)	1,71%	9 871 (1,60%)	1,68%
Juh	8	9 674	87 (0,01%)	1,58%	1 763 (0,29%)	1,85%
	50	3 284	18 354 (2,99%)	3,50%	27 639 (4,50%)	5,01%

Forrás: KSH adatbázis alapján saját számítás

$$\sum_{h=1}^7 n_h$$

$$\sum_{h=8}^{14} n_h \quad \text{és zárójelben: } \left( \frac{\sum_{h=8}^{14} n_h}{\sum_{h=8}^{14} N_h} \right)$$

A két utóbbi táblázatot összevetve azt tapasztaljuk, hogy az optimumot adó réteghatárok módosulnak, kivéve a juhok esetében. Kiemelendő különbség még az is, hogy bár állatfajonként nem módosult, hogy a rétegzésen belül melyik elosztás adja a kisebb elemszámot, a két elosztás közötti nagyságrendi eltérések jelentősen megváltoztak. A szarvasmarháknál az OPS mellett négyszeres mintára lenne szükség Neyman-féle optimális allokációval az arányoshoz képest. Újra hangsúlyozom, hogy csakúgy, mint az előző fejezetben, itt is csak a „kisebb” gazdaságok rétegeiben lehetett igazán optimalizálni az allokációt a szórások felhasználásával, mivel a „nagyobbak” rétegeiből teljes körű kiválasztást kellett eszközölni. A sertések és juhok esetében matematikailag még mindig jobbnak, hatékonyabbnak bizonyul a Neyman-féle elosztás, de a különbségek meglehetősen lecsökkentek, mindössze néhány tized százalékra. Ez elgondolkodtató, hogy a sokkal egyszerűbb, és kevesebb előzetes információra és számításra épülő arányos elosztást mégis megéri alkalmazni az elenyésző különbségek miatt.

Az előző fejezetben alkalmazott eljárásnak megfelelően a teljes sokaságra is kiszámíthatjuk a szükséges minta nagyságát, illetve arányát. 15 785 a „nagyok” rétegének maximális értéke. Ehhez a „kisebb” gazdaságokból további 4 433 kiválasztására van szükség.

Összesen tehát  $\frac{20218}{617500} = 3,27\%$  a szükséges kiválasztási arány. A KSH-ban használt minta

2007-ben érvényes nagysága (40 000) ehhez képest gyakorlatilag a duplája, hiszen a SEFF mutató értéke 0,506. Ezzel a mintaelemszám felét megspórolhatjuk, ami még a várakozásaimat is felülmúlóan jó eredmény.

Természetesen nem szabad figyelmen kívül hagyni azt a fontos tény, hogy az általam kidolgozott és javasolt mintavételi terv egy földrajzilag meglehetősen szétszórt mintasokaság kiválasztását eredményezi. Ez pedig az egy kérdőívre jutó összeírási költségek akár lényeges növekedését is okozhatja. Ugyanakkor a mintanagyság közel felére csökkenése ugyanolyan költségvetési keretek mellett is megengedi az egy adatlapra eső kiadás növelését. (Nem a duplázódását, mert a település, illetve körzetfelkeresési díjak nem állnak egyenes arányban a kérdőívek számával.) Azt sem szabad elfelejteni, hogy ezzel „csak” az állatállomány kérdésében jutottunk új eredményhez. A mezőgazdaság statisztikának viszont a földterülettel kapcsolatban is biztosítani kell a megfelelő minőségű adatokat. A leginkább hangsúlyozandó „vívmány” viszont az, hogy ez **a kisebb, és talán valamivel drágább minta megfelelő minőségű adatokat biztosít, amelyek megfelelnek az EU-s jogszabályok által támasztott követelményeknek.**

A dolgozat első felében már szót ejtettem arról, hogy a csoportos mintavétellel történő szakítás abba az irányba mutat, hogy csakis egy jóval alacsonyabb elemszámú minta lehet hatékony a magas összeírási költségek miatt. A gyakorlati mintavételes összeírásokban jártas statisztikusok számára, így számomra is, az 50 százalékos mintaelemszám csökkentés, úgy gondolom már önmagában is egy rendkívül szép eredménynek mondható. Azt hozzátéve, hogy minőségileg is jobb, illetve megfelelő becsléseket eredményez, csak emeli az elért eredmény értékét.

#### ***5.4. Kísérleti eredmények a 2010-es Általános Mezőgazdasági Összeírásból***

A KSH hozzájárulásával és segítségével lehetőségem nyílt a 2010-es census (ÁMÖ2010) adatain is kipróbálni a fentiekben kidolgozott és javasolt mintavételi tervet. Azért nagyon jó lehetőség ez, mert minden gazdaság adata rendelkezésre áll, és azon túl, hogy számszerűsítem a mintavételi hibát, illetve kiszámítom a paraméterek pontbecslését, egyben össze is tudom hasonlítani a tényleges értékösszegekkel a kapott eredményeket országos és regionális szinten egyaránt. Így tehát nem csak a számított standard hibák mentén van lehetőségem megítélni a mintavétel hatékonyságát, hanem a valósággal is összevethetőek a számok. Ilyen alkalom pedig csak kivételes esetekben nyílik a statisztikus számára, ezért külön öröm számomra, hogy hozzáférhetővé tették a 2010-es ÁMÖ vonatkozó adatait is a kutatás teljesebbé tétele érdekében.

A 48. táblázatban kiemelt réteghatároknak megfelelően, a 2007-es adatok alapján végzett kalkulációk alapján azt találtuk, hogy – az átfedéseket nem számolva – 15785 kiemelt gazdaságot kellene teljes körűen kiválasztani. Az ÁMÖ2010 adatai alapján 14989 azon egyéni gazdaságok száma, akik legalább az egyik állatfaj szempontjából elérik, illetve meghaladják a réteghatárnak választott értéket. Összességében ez egészen jó közelítésnek tűnik. Ugyanakkor árnyalja a képet, ha állatfajonként megvizsgáljuk, hányan számítanak bele a kiemelt körbe 1-1 állatfaj „miatt”, és ezt hasonlítjuk össze a korábbi „becslésekkel” (lásd 49. táblázat).



49. TÁBLÁZAT: KIEMELT GAZDASÁGOK BECSÜLT ÉS TÉNYLEGES SZÁMA

Állatfaj	Becsült gazdaságszám	Tényleges gazdaságszám
Szarvasmarha	1 829	2 751
Sertés	4 282	2 982
Juh	9 674	10 335

*Forrás: KSH adatbázis alapján saját számítás*

Összehasonlítva a tényleges, és becsült gazdaságszámokat azt láthatjuk, hogy legjobban a juhok szerint nagynak minősülőket sikerült közelíteni. Ebből arra következtethetünk, hogy a juhtartók körében történt a legkisebb változás, elmozdulás. A másik két állatfajnál egészen mást vehetünk észre. Az általam – mintavételi szempontból – nagynak minősített szarvasmarha-tartók száma másfélszerese a 2007-es adatokból számítottaknak. Ez jelenthet egyfajta koncentrációs folyamatot, hiszen a szarvasmarha állomány növekedése nem kíséri a jelenséget. A sertések esetében éppen fordított a helyzet. Az vizsgált 3 év alatt  $\frac{3}{4}$ -ére esett vissza a nagyobb gazdaságok száma, és ezzel egyidejűleg az állomány országos nagysága is csökkenést mutat. Ezek alapján sok egyéni gazdaság visszaszorította, vagy teljesen abbahagyta a sertéstenyésztést. Persze azt is érdemes szem előtt tartani, hogy a gazdaságok száma 2007 és 2010 között még mindig közel 10 százalékkal zsugorodott. A 2.4.1 fejezetben bemutatott folyamat tehát továbbra is él, és a gazdaságszám visszaesése mellett strukturális változások is végbemennek.

A réteghatár alatti gazdaságokból az ajánlott mintanagyság 4433. Így, a kiemelttekkel együtt összesen 19422 gazdaság került a mintába az ÁMÖ2010-es adatait felhasználva. Ez, a tényleges gazdaságszámhoz viszonyítva 3,4 százalékos mintavételi arányt jelent.

A mintavételezést, mivel a teljes adatállomány rendelkezésemre állt a három állatfaj adatait tekintve, tetszőleges ismétlésszámmal el lehet végezni. A konkrét értékek bemutatása során egy realizálódott minta alapján számított értékeket fogok bemutatni. Azt tartom fontosnak hangsúlyozni, hogy a véletlen kiválasztást sokszor megismételve szépen érvényesült az a valószínűségelméleti háttérrel alátámasztott állítás, hogy a mintából becsült értékösszeg a tényleges állatállomány létszám körül ingadozik, és a megbízhatósági szintnek megfelelően,

az intervallumbecslés határai közé esik a valós érték. Ezt a véletlen kiválasztás biztosítja. Nem tapasztalható egyik irányban sem szisztematikus torzítás a becslésekben.

50. TÁBLÁZAT: RÉGIÓS BECSLÉSEK EREDMÉNYEI (EGYÉNI GAZDASÁGOK)<sup>43</sup>

Állatfaj/Régió		Közép-Magyarország	Közép-Dunántúl	Nyugat-Dunántúl	Dél-Dunántúl	Észak-Magyarország	Észak-Alföld	Dél-Alföld
Szarvasmarha	ÁMÖ	15 468	18 628	26 867	20 072	24 304	66 692	71 224
	Becslés	14 083	19 014	32 923	21 184	24 942	55 249	71 859
	CV	16,6%	6,9%	11,3%	11,8%	9,3%	3,9%	9,3%
Sertés	ÁMÖ	58 223	99 703	119 662	114 778	59 936	202 236	322 215
	Becslés	64 165	92 007	123 737	99 705	55 869	208 107	308 266
	CV	13,2%	2,9%	5,8%	5,3%	10,0%	4,2%	6,1%
Juh	ÁMÖ	56 473	80 050	24 449	96 389	87 118	381 417	316 437
	Becslés	54 843	78 830	24 168	97 568	87 108	381 774	331 164
	CV	2,8%	1,0%	4,1%	1,9%	1,6%	0,7%	2,2%

*Forrás: KSH adatbázis alapján saját számítás*

Ugyan a régiókra vonatkozóan nem támaszt elénk az EU konkrét követelményeket a hibaszámokkal kapcsolatban, mivel ez a területi szint fontos a hazai publikációk szempontjából, az 50. táblázatba összefoglaltam a régiós becslések eredményeit az egyéni gazdaságokra. A színezett cellák azt jelentik, hogy 68 százalékos megbízhatóság mellett az intervallumbecslés a megvalósult mintában tartalmazza a tényleges értéket. A relatív standard hibák sorai a leginkább érdekesek. A juhok esetében meglehetősen jó minőségű becslésekről beszélhetünk (talán Nyugat-Dunántúl kivételével), hiszen az értékek nem haladják meg a 3 százalékot. A sertések és a szarvasmarhák vonatkozásában már nem lehetünk ennyire elégedettek, hiszen akár a 10 százalékot elérő értékek sem ritkák. (Ne felejtjük el, hogy a „szokásos” 95 százalékos megbízhatósági szint mellett 1,96-tal kellene szorozni a hibaszámokat!) A sertések és szarvasmarhák becsült értékösszeg adatait tehát egy ilyen mintavételi terv mellett óvatosan kell kezelni, és mindenképpen fel kell tüntetni a mintavételi hiba mértékét, hogy az esetleges felhasználók annak tudatában értékelhessék az adatok minőségét.

<sup>43</sup> ÁMÖ: ÁMÖ2010 adata, CV: relatív standard hiba (68 százalékos megbízhatósági szint mellett)

51. TÁBLÁZAT: ORSZÁGOS SZINTŰ BECSLÉSEK EREDMÉNYEI (EGYÉNI GAZDASÁGOK)

Állatfaj	ÁMÖ2010	Becsült érték	Relatív standard hiba (CV)
Szarvasmarha	243 255	239 254	3,7%
Sertés	976 753	951 856	2,6%
Juh	1 042 333	1 055 456	0,8%

Forrás: KSH adatbázis alapján saját számítás

A dolgozat célja elsősorban azonban nem a régiós szintű jó minőségű adatok elérése volt, hanem az EU-s jogszabályokban meghatározott, az ország teljes állományára vonatkozó hibaszámoknak való megfelelés<sup>44</sup>. Az 51. táblázat a „próba” mintavétellel elért eredményeket tartalmazza. A 15. táblázatban található, egyéni gazdaságokra kiszámított határértékeknek kell megfelelni (szarvasmarha, sertés és juh sorrendben: 16,04%, 6,18% és 2,35%). **Összehasonlítva a relatív standard hiba értékeket, azt látjuk, hogy bőven az elvárások alatt maradnak a számok.** Olyannyira, hogy valószínűleg jóval **kisebb mintanagyság is elegendő volna** a kissé megváltozott 2010-es mezőgazdasági alapsokaság állatállomány adatainak felméréséhez. A réteghatárok módosításával, illetve a „kisebb” gazdaságok köréből szükséges mintanagyság módosításával tovább csökkenthető az összeírásra kijelölt sokasági elemek száma. Ebből számomra két fontos következtetés vonható le:

- A korábban **javasolt mintavételi módszer valóban hatékonyabb becsléseket eredményez** a jelenleg alkalmazotthoz képest, vagyis talán tényleg ez lehet a jó irány.
- A magyar mezőgazdaságban még mindig olyan **gyors ütemű változások zajlanak**, hogy a mintavételi terv kialakításához, illetve a rétegzéshez már akár a néhány évvel korábbi információ sem ad pontos alapot. Célszerű tehát folyamatosan, akár az évközi rendszeres összeírások adatai alapján „alkalmazkodni” a változó struktúrához.

<sup>44</sup> A becslések és a hibaszámok az egyéni gazdaságokra vonatkoznak. Ugyanakkor a gazdasági szervezetek teljes körűen begyűjtött adatai természetesen mérséklék a mintavételi hiba relatív nagyságát a teljes régiós, illetve országos állományokra vonatkozóan.

## 6. Összegzés

A mezőgazdaság termelésstatisztikai rendszerének adatgyűjtései rendkívül szerteágazó, összetett és bonyolult folyamatot jelentenek. Az évtizedek során folyamatosan bővülő adatigény és a statisztika módszertan, illetve a számítástechnika fejlődése egyre pontosabb adatokat eredményező, ugyanakkor jóval bonyolultabb mintavételi módszerek alkalmazását tette lehetővé a hivatalos adatgyűjtések során. Magyarországon a gazdaságok elaprózódása, és rendkívüli számossága sajátos jellemzőket alakított ki. A mintavételi technikáknak ezekhez a feltételekhez kell igazodnia a jogszabályokban meghatározott pontossági követelmények betartása mellett.

A dolgozat egy statisztika-elméleti problémára (hatékonyságjavítás lehetőségei erősen aszimmetrikus eloszlású sokaságokból vett mintákra) keres megoldást, úgy, hogy közben konkrét gyakorlati példán vezeti végig a számításokat, a magyarországi mezőgazdaság, illetve az állatállomány összeírások kapcsán. A kutatás végeredményeképpen a disszertációban 6 tézis került kimondásra a bevezetőben megfogalmazott 6 hipotézis „tesztelése” után. Ezek közül 4 inkább módszertani jellegű tézis, 2 pedig konkrétan az állatállomány, illetve a mezőgazdaság sajátosságain alapuló gyakorlati megállapítás, új eredmény. Összegzésképpen tehát a kutatás során feltárt újszerű eredményeket emelném ki, a tézisek felelevenítésével és összefoglalásával.

Statisztika-módszertani szempontból véleményem szerint a legfontosabb eredménye a dolgozatnak a rétegzett mintavétel csoportosításának kialakítása. A rétegzés a statisztika oktatásban, és sokszor a módszertannal foglalkozó szakirodalomban is egy hatékonyságjavító technikaként terjedt el. Ugyanakkor a rétegzést meghatározó okok, motivációk sokszor csak a rétegzés technikáját használják fel, anélkül, hogy érdemben javítanák a mintavételi eredményekből készített becslések hatékonyságát. Ezért tartom fontosnak elkülöníteni a technikai rétegzést attól a kapcsolati rétegzéstől, ami kifejezetten a vizsgált változóval, vagy akár változókkal lehetőség szerint erős sztochasztikus kapcsolatban álló réteggépző ismerv megválasztásával és felhasználásával jut el hatékonyabb becslésekhez.

**1. tézis:** T1/A: A rétegzett mintavételt, mint gyakran alkalmazott mintavételi eszközt, a réteggépző ismerv tulajdonságai, illetve a rétegzés okai, céljai szerint két alcsoportra

bonthatjuk: kapcsolati és technikai rétegzésre. T1/B: A kapcsolati rétegzés célja a becslések hatékonyságának javítása, míg technikai rétegzés esetén a rétegzés módszerét, csak mint osztályozó eszközt használjuk fel, de a becslések hatékonysága nem feltétlenül javul.

A statisztikai adatok minőségére vonatkozó konkrét előírások általában pontosan meghatározzák, hogy mi az abszolút vagy relatív hibahatár, ami még elfogadható. Így van ez a mezőgazdasági adatokra vonatkozó jogszabályokban is. Ehhez képest a mintavételi módszerek hatékonyságát összehasonlító szemléletben azonos mintanagyság mellett szokás megvizsgálni és összevetni a becslések hatékonyságát. A gyakorlat szempontjából viszont célszerűbbnek tűnt a rögzített, előírt hibahatárt biztosító mintanagyságokat összehasonlítani különböző mintavételi tervek mellett. Az úgynevezett mérethatásosság mutató (size efficiency - SEFF) a gyakorlati munka során jobban alkalmazható mérőszámnak bizonyult.

**2. tézis:** Az úgynevezett mérethatásosság mutató (SEFF) a gyakorlati munka során a tervhatásossági mutatónál (DEFF) jobban alkalmazható mérőszám a különböző mintavételi tervek összehasonlítására, hiszen ekkor adott standard hiba mellett hasonlítjuk össze a szükséges mintanagyságokat.

A mezőgazdaság statisztika esetében is bizonyítást nyert az a tény, miszerint a régió – bár kötelező szempont a rétegzés kialakításánál –, mint réteggépző ismerv egyáltalán nem, illetve jelentősen nem javítja a becslések hatékonyságát. Ezek alapján tehát csak technikai rétegzésről beszélhetünk, amikor a gazdaságok régió szerinti hovatartozását felhasználjuk a mintavételi terv kialakítása során. Ugyanakkor a régiók szerinti rétegzés a területi reprezentativitás miatt sok esetben feltétlenül szükséges.

**3. tézis:** A régió, mint réteggépző ismerv csak a Neyman-féle optimális allokáció felhasználásával hoz javulást az egyszerű véletlen mintavételhez képest, és a javulás mértéke ugyan állatfajonként különböző, de sehol sem nevezhető jelentősnek.

Többször előkerül a dolgozatban, hogy a mezőgazdaságban érdekelt egyéni gazdaságok jellemzői sok tekintetben szélsőségesnek nevezhető tulajdonságokkal bírnak. Így van ez az

állatállományok szempontjából is. Közepes, illetve erős jobb oldali aszimmetriát mutat az eloszlásuk. A szakirodalom szerint ezt a tényt körültekintően kell kezelni, mert torzíthatja a mintavételből származó becsléseket, ha figyelmen kívül hagyjuk a sokaság ilyen jellemzőjét.

A kevésbé aszimmetrikus eloszlást mutató sertés és szarvasmarha esetében, ha két nagyságkategória kialakításában gondolkodunk, akkor a réteghatároló pontot nem szükséges teljesen pontosan meghatározni, mert egy viszonylag tágabb intervallum mellett is nagyon hasonlóan alacsony mintaelemszám szükséges a megcélzott hibahatár eléréséhez. Ez gyakorlati szempontból egy nagyon hasznos észrevétel, hiszen tompítja azoknak az esetlegesen pontatlanul, régebből rendelkezésre álló szórásadatoknak a torzító hatását, amelyeket felhasználunk az optimális réteghatár meghatározásakor.

**4. tézis:** Kisebb terjedelemmel rendelkező eloszlások esetében a nagyságkategóriák szerinti kapcsolati rétegzésnél a minimálisan szükséges jelentősen nem meghaladó mintaelemszámot eredményező mintavételi tervekhez tarozó réteghatár nem egy pontba, hanem egy nagyobb tartományba esik, ami megkönnyíti a meghatározását.

A mintavételi tervek kialakításával foglalkozó statisztikusok számára gyakran nem csak a mintavétel módszerének meghatározása okoz fejtörést. A gyakorlatban sokszor előfordul, hogy maga az alapsokaság meghatározása, megfelelő, célszerű definiálása is problémákat vet fel, és a kompromisszumokat már ezen a ponton is meg kell kötni. Az alapsokaság lehatárolása a mezőgazdaság statisztikában a gazdasági szervezeteknél a végzett tevékenység alapján történik, az egyéni gazdaságok esetében pedig egy megfelelően meghatározott gazdaságküszöb elérése jelenti az alapsokasági tagságot. Már a dolgozat bevezetőjében felvettem, hogy a mezőgazdasági termelés két fő iránya szerint (növénytermesztő illetve állattartó gazdaság) érdemes lehet elkülöníteni az egyéni gazdaságokat akkor, amikor különböző témájú adatgyűjtésekhez jelöljük ki őket. Azt gondoltam, hogy ha az állatok megfigyeléséhez csak a ténylegesen állatok tartásával foglalkozó (és nem tisztán növénytermesztő) gazdaságokat tekintjük alapsokaságnak, akkor azzal már a mintavétel előtt egyfajta homogenizálást hajtunk végre, és ezért jobb kiinduló helyzetből vágunk neki a mintavételi terv kialakításának, így az hatékonyabb lesz. Bármennyire is tűnik ez az elképzelés logikusnak, a szimulációs számítások meggyőzően nem támasztották alá.

**5. tézis:** Nem eredményez alacsonyabb kiválasztási arányokat az alapsokasági szemlélet szűkítése, és csak az állattartó gazdaságok figyelembe vétele.

Az úgynevezett „take all – take some” mintavételi elvet kifejezetten jobb oldali aszimmetriával rendelkező eloszlások mellett javasolják. Ennek lényege, hogy a „nagyobb” egyedek körét (ami viszonylag kis számosságot jelent) teljes egészében megfigyeljük (take all), és a „kisebbeket” pedig (akik viszonylag sokan vannak) mintasokaságon keresztül (take some) írjuk csak össze. Mindhárom állatfaj vonatkozásában megvizsgáltam ezt a lehetőséget is. Bár az aszimmetrikus eloszlás mindre jellemző, a Neyman-féle optimális elosztáshoz képest csak a juhok esetében kaptam jobb eredményeket a fent vázolt módszer alkalmazásával. Összehasonlítva a három állatfaj jellemzőit, arra a következtetésre jutottam, hogy a „take all – take some” technika alkalmazásának hatékonyságjavítása két dologtól függ: egyrészt az aszimmetria mértékétől, másrészt a változó értékeinek nagyságrendjétől, terjedelmétől.

**6. tézis:** A „take all – take some” mintavételi technika és a Neyman-féle optimális elosztás (2 réteget feltételezve) egymáshoz viszonyított hatékonyságát a vizsgált változó aszimmetriája és terjedelme határozza meg.

A dolgozat végére kialakított mintavételi terv gyakorlatilag felére csökkenti a szükséges mintavételi arányt az állatállomány megfigyelésekhez. Bár a területi szétszórtság nem jelent egyben költséghatékonyabb megoldást, de mégis jóval kisebb adatszolgáltatói terhek mellett lehet megfelelő minőségű becsléseket készíteni a kiemelt állatfajokra, teljesítve ezzel az EU-s jogszabályi előírásokat. A javasolt módszer 2010-es adatokon történő „tesztelése” pedig azt vetítette előre, hogy még ennél alacsonyabb mintaelemszámmal is elérhető lenne az elvárt minőség a megváltozott sokasági jellemzők alapján.

Dióhéjban ezek a kutatás során elért eredmények. Talán nem túlzás kijelenteni, hogy az összes megfogalmazott tézisnek van gyakorlati vetülete, és előrébb viheti a mezőgazdaság statisztikában alkalmazott mintavételi módszerek fejlesztését. 5 éve, amikor először felmerült bennem az önálló kutatás gondolata, akkor is ilyen motivációk vezéreltek. Talán sikerült is

megvalósítani az eredeti célkitűzéseket, és mellette még néhány elméleti jellegű újszerű megközelítés is megszületett a munka során.

### **További kutatási lehetőségek**

A dolgozatból látható, hogy az eddigi munka is meglehetősen sok számítást és időt igénylő folyamat volt. Megállni azonban még mindig nem szabad. Vannak még kiaknázható lehetőségek, amelyek már ekkora hozadékkal talán nem kecsegtetnek, de még mindig javítani tudják a mintavétel hatékonyságát, vagyis a szükséges mintavételi arányt. Potenciálisan három további kutatási irányt látok még a témában:

1. A két rétegzési szempont (régiók szerinti technikai és nagyságkategóriák szerinti kapcsolati rétegzés) kombinálása során az is világossá vált, hogy még hatékonyabb eredmények érhetőek el, ha régióként eltérő réteghatárokat húzunk meg a két réteg között. Ez bonyolítaná ugyan a mintavételi tervet, de a mai számítástechnikai háttér mellett ez nem jelenthet akadály. Az is igaz, hogy ez további, idő és munkaigényes számításokat igényel, de a megfelelő hozadék reményében ezeket sem lenne haszontalan kivitelezni.
2. Eddig nem vizsgáltam meg az utólagos rétegzés lehetőségét, ami szintén javíthatná az eredményeket. A módszer lényege, hogy a rétegzést a mintavétel után végezzük el. Ebből több előny is származhat. Egyrészt ez a technika védelmet jelent a kedvezőtlen mintabeli eloszlások kialakulása ellen. Másrészt mintavétel után a rétegző ismérveket változtatni tudjuk, különbözőképpen határozhatjuk meg az éppen becsülni kívánt változókhoz igazodóan, vagyis többféle rétegzést is kialakíthatunk. Ez utóbbi tulajdonság a többcélú felvételek esetén rendkívüli jelentőséggel bír. (Holt-Smith 1979; Éltető 1970; Éltető 1982.)
3. Az utólagos rétegzéshez kiegészítésképpen meg lehetne vizsgálni részletesen, hogy a klaszteranalízis segítségével mennyivel tudnánk javítani a mintavétel, illetve a becslések hatékonyságát. (Hajdu 2003) Ez a módszer egyszerre több ismerv szerint alakítja ki a lehető leghomogénebb csoportokat. A bevont változók számától és a megcélzott csoportszámától függően a homogenitás szintje természetesen változik, de a meglévő adatokból a legjobb megoldás érhető el. Ezt előzetes rétegzéshez azért nem



lehet felhasználni, mert a klaszteranalízis eredménye ugyan minden egyedet besorol egy csoportba, de a csoportba sorolás „szabályait” visszafelé nem lehet megadni, vagyis ezzel nem lehet meghatározni a konkrét rétegzési elveket. Így csak utólagos rétegzéshez adhat támogatást a módszer. A lehetőség azonban így is adott. Egyszerre több, fontos változót felhasználva így lehet a leghatékonyabb módon csökkenteni a csoporton belüli szórások mértékét, nagyságát, ami a becslések hatékonyságának javulását eredményezi. (A témában lásd: Galambosné Tiszberger 2011b) Kérdés persze, hogy melyek a legfontosabb változók, és mennyi az optimálisnak nevezhető csoportszám. A további kutatások során ezekre a kérdésekre is választ kellene keresni.

## Ábrák jegyzéke

1. ábra: Az 1895-ös census kiadványának címlapja .....	18
2. ábra: Az 1895-ös census kérdőívének egyik állatállománnyal foglalkozó táblája .....	23
3. ábra: A 2007-es gszo egyik táblája.....	24
4. ábra: 3 állatfaj állományának alakulása Magyarországon, 1895-2010 (1895=100%) .....	27
5. ábra: 3 állatfaj állományának alakulása Magyarországon, 1970-2010 (december 1.) (1970=100%) .....	29
6. ábra: A földterület alakulása főbb művelési ágak szerint, Magyarországon, 1950-2010 (1950=100%)..	30
7. ábra: Az egyéni gazdaságok SFH értékének megoszlása a gazdálkodás célja szerint, 2003-2005-2007 ..	39
8. ábra: Az egyéni gazdaságok számának megoszlása a gazdálkodás célja szerint, 2003-2005-2007 .....	40
9. ábra: A gazdaságok számának alakulása, 1972-2010 (1972=100%) .....	42
10. ábra: A mintavételi módszerek csoportosítása .....	64
11. ábra: Sertést, juhot, illetve szarvasmarhát tartó gazdaságok .....	82
12. ábra: Az állatállomány szezonális ingadozása az egyéni gazdaságokban (ezer db) .....	83
13. ábra: Az állatállomány szezonális változásai az egyéni gazdaságokban (előző időszak = 100%) .....	84
14. ábra: A szarvasmarha állomány hisztogramja.....	94
15. ábra: állatállomány arányok halmazok szerint .....	99
16. ábra: Lorenz-görbék.....	100
17. ábra: Az egyéni gazdaságok átlagos SFH értékeinek alakulása régióként, 2007 (€).....	117
18. ábra: A gazdasági szervezetek átlagos SFH értékeinek alakulása régióként, 2007 (€) .....	118

## Táblázatok jegyzéke

1. táblázat: A gazdaságküszöb alakulása a teljes körű összeírások során .....	38
2. táblázat: A gazdaságok számának és SFH értékének alakulása, 1972–2010 .....	41
3. táblázat: Az egyéni gazdaságok és a gazdasági szervezetek arányeltolódásai, 2000–2007 .....	43
4. táblázat: A rendszeres mezőgazdasági összeírások .....	45
5. táblázat: Az aszimmetria alakulása állatfajonként, különböző mutatókkal .....	90
6. táblázat: Szarvasmarha-állomány osztályközös gyakorisági sora .....	95
7. táblázat: A szarvasmarha-állomány leíró statisztikája különböző sokaságok szerint .....	95
8. táblázat: Sertésállomány osztályközös gyakorisági sora .....	96
9. táblázat: A sertésállomány leíró statisztikája különböző sokaságok szerint.....	96
10. táblázat: Juhállomány osztályközös gyakorisági sora .....	97
11. táblázat: A juhállomány leíró statisztikája különböző sokaságok szerint.....	98
12. táblázat: Maximális relatív standard hiba állatfajonként az EU előírások alapján .....	101
13. táblázat: A kiemelt állatfajok jellemző adatai, 2007.....	102
14. táblázat: Megengedhető maximális értékösszegre vonatkozó hibanagyságok az egyéni gazdaságoknál .....	104
15. táblázat: Megengedhető maximális relatív hibanagyságok az egyéni gazdaságokra vonatkozóan (%) .....	104
16. táblázat: Egyes korábbi időszakok hibaszámai .....	105
17. táblázat: A relatív szórás alakulása sokaságok és fajok szerint (%) .....	106
18. táblázat: Állatfajonként szükséges EV mintavételi arány, különböző szemléletű alapsokaság mellett (%) .....	109
19. táblázat: Szükséges mintaválasztási arányok a teljes alapsokasághoz viszonyítva (%) .....	111
20. táblázat: A jelenleg alkalmazott rétegzés szerint az egyéni gazdaságok megoszlás a 2007-es adatok alapján (%) .....	113
21. táblázat: A jelenleg alkalmazott rétegzés szerint az egyéni gazdaságok SFH értékének megoszlása a 2007-es adatok alapján (%) .....	113
22. táblázat: A rétegzés változása időben .....	114
23. táblázat: A szükséges mintavételi arányok EV, illetve rétegzés mellett (%) .....	119
24. táblázat: Szórásfelbontás az egyéni gazdaságoknál régiós csoportosítás mellett .....	120
25. táblázat: Szórásfelbontás az állattartó egyéni gazdaságoknál régiós csoportosítás mellett .....	121
26. táblázat: Az egyes állatfajok szórása, és az egyéni gazdaságok száma régióként .....	121
27. táblázat: Az egyes állatfajok szórása, és az állattartó egyéni gazdaságok száma régióként .....	121
28. táblázat: A szükséges mintavételi arányok különböző mintavételi tervek esetén .....	123
29. táblázat: Különböző régiós rétegzés és az EV mintavételi terv összehasonlítása .....	126
30. táblázat: A „kisebb” gazdaságok aránya, és a hozzájuk tartozó állatállományi arányok, 2007.....	128

31. táblázat: Optimális réteghatár arányos allokáció mellett .....	131
32. táblázat: SEFF arányos nagyság szerinti rétegzés és ev mintavétel mellett .....	131
33. táblázat: Szükséges mintanagyság különböző réteghatárok mellett (Neyman-féle optimális) .....	133
34. táblázat: SEFF különböző réteghatárok mellett.....	136
35. táblázat: Szóráshányados értékek különböző réteghatárok mellett.....	137
36. táblázat: Szükséges mintanagyság különböző réteghatárok mellett (take all – take some) .....	141
37. táblázat: SEFF különböző réteghatárok mellett (TATS/EV, TATS/AR).....	142
38. táblázat: SEFF különböző réteghatárok mellett (NO/TATS).....	143
39. táblázat: Legalacsonyabb szükséges mintaelemszámok összehasonlítása.....	143
40. táblázat: Szarvasmarhát tartó gazdaságok gyakorisági sora .....	146
41. táblázat: Sertést tartó gazdaságok gyakorisági sora .....	146
42. táblázat: Juhot tartó gazdaságok gyakorisági sora .....	146
43. táblázat: Ekman-szabály számítása a szarvasmarháknál .....	148
44. táblázat: Ekman-szabály számítása a sertéseknél .....	148
45. táblázat: Ekman-szabály számítása a juhoknál.....	149
46. táblázat: Az exponenciális illesztés főbb eredményei .....	152
47. táblázat: Szükséges mintanagyság régiós és nagyságkategóriás rétegzés mellett .....	155
48. táblázat: Szükséges mintanagyság régiós és nagyságkategóriás rétegzés mellett a valós alapsokaságra becsülve (N=618 500) .....	158
49. táblázat: Kiemelt gazdaságok becsült és tényleges száma .....	161
50. táblázat: Régiós becslések eredményei (egyéni gazdaságok) .....	162
51. táblázat: Országos szintű becslések eredményei (egyéni gazdaságok) .....	163

## Irodalomjegyzék

- [1.] **Ay, János – Oros, Iván:** *Reprezentatív módszerek alkalmazása a gyümölcstermelési statisztikában* = Statisztikai Szemle 45. évf. 1967. 6. sz. 675-684. p.
- [2.] **Ay, János:** *A mintavételes állatösszeírások módszertani kérdései*. Kandidátusi értekezés. Budapest. 1976. 150 p.
- [3.] **Benoist, György – Falder-Huerta, A.:** *A szőlő- és gyümölcsösültetvények statisztikai megfigyelése az EU-ban* = Statisztikai Szemle 80. évf. 2002. 7. sz. 637-647. p.
- [4.] **Bertáné Németh, Ágnes:** *Az irodalomkutatástól a hivatkozásig*. Pécs. Pécsi Tudományegyetem Közgazdaságtudományi Kar, 2001. 42 p.
- [5.] **Buday-Sántha, Attila:** *A magyar agrár- és vidékfejlesztés ellentmondásai*. Magyar Tudomány, 2009. 170. évf. 8. sz. 937-945. p.
- [6.] **Buday-Sántha, Attila:** *Agrár- és vidékpolitika*. Budapest: SALDO Kiadó. 2011. 377 p.
- [7.] **Cochran, William G.:** *Comparison of Methods for Determining Stratum Boundaries*. = Bulletin of the International Statistical Institute. 38. évf. 1961. 2. sz. 345-358. p.
- [8.] **Cochran, William G.:** *Sampling Techniques*. New York: John Wiley & Sons, INC. 1977. 428 p.
- [9.] **Dalenius, T.:** *The Problem of Optimum Stratification*. = Skand Aktuarietidskr. 33. 1950. 203-213. p.
- [10.] **Dalenius, T. – Hodges, J.L.:** *Minimum Variance Stratification*. = Journal of the American Statistical Association. 54. 1959. 88-101. p.
- [11.] **David, I. P.:** *Sampling Strategy for Agriculture Censuses and Surveys in Developing Countries*. Washington, International Congerence on Agriculture Statistics 2000. 18-20 March 1998. 15 p.
- [12.] **Ekman, G.:** *An Approximation Useful in Univariate Stratification*. = The Annals of Mathematical Statistics. 30. 1959. 219-229. p.

- [13.] **Elsayed**, Ali Habib: *Estimation of modified measure of skewness* = Electronic Journal of Applied Statistical Analysis. Vol. 4. Issue 1. 2011. 56-70. p.
- [14.] **Éltető**, Ödön: *Mintavétel véges alapsokaságból*. MTESZ Bolyai János Matematikai Társulat. Budapest. 1970. 154 p.
- [15.] **Éltető**, Ödön: *Mintavételi eljárások*. In: Éltető Ö. – Meszéna Gy. – Ziermann M.: Sztochasztikus módszerek és modellek, Közgazdasági és Jogi Könyvkiadó, Budapest. 1982. 420 p.
- [16.] **Fraller**, Gergely: *Szemelvények a mintavételi rétegzés területéről* = Statisztikai Szemle 89. évf. 2011. 4. sz. 357-378. o.
- [17.] **Farver**, Thomas B.: *Some Practical Considerations in Sampling Livestock Populations to Estimate Disease Prevalence and Other Parameters* = Preventive Veterinary Medicine 2. 1984. 453-462. p.
- [18.] **Galambosné Tiszberger**, Mónika: *Kisgazdaságok a magyar mezőgazdaságban*. Pécs. Pécsi Tudományegyetem Közgazdaságtudományi Kar Regionális Politika és Gazdaságtan Doktori Iskola Évkönyve, 2009. 60-69. p.
- [19.] **Galambosné Tiszberger**, Mónika: *Megyei és regionális szintű adatok a mezőgazdaság statisztikában*. Pécs. „Félidőben” IV. Országos Környezetgazdaságtani PHD-Konferencia, 2. kötet, 2010. október 8-9. 164-174. p.
- [20.] **Galambosné Tiszberger**, Mónika: *A rétegzett mintavételről* = Statisztikai Szemle 2011a 89. évf. 9. sz. 909-929. p.
- [21.] **Galambosné Tiszberger**, Mónika: *Rétegzési lehetőségek az állatállomány összeírásoknál*. In: Lázár, Ede (szerk.): Gazdasági és üzleti kihívások a Kárpát-medencében című konferencia kiadványkötete. Sapientia Egyetem. Státus Kiadó, Csíkszereda, 2011b 132-142. p.
- [22.] **Gömbös**, Ervin: *Magyarok az ENSZ apparátusában* = Külügyi Szemle 7. évf. 2008. 2. sz. 213-229. p.
- [23.] **Hajdu**, Ottó: *Többváltozós statisztikai számítások*. Budapest. Központi Statisztikai Hivatal, 2003. 457 p.

- [24.] **Hajdu, Ottó – Pintér, József – Rappai, Gábor – Rédey, Katalin:** *Statisztika I.* Pécs. Janus Pannonius Tudományegyetem Közgazdaságtudományi Kar, 1994. 352 p.
- [25.] **Hedlin, Dan:** *A Procedure for Stratification by an Extended Ekman Rule* = Journal of Official Statistics. 2000. Vol. 16. No. 1. 15-29. p.
- [26.] **Heilbron, David C.:** *Zero-altered and other regression models for count data with added zeros.* = Biometrical Journal. 1994. Vol. 36. Issue 5. 531-547. p.
- [27.] **Hidiroglou, M. A.:** *The Construction of a Self-representing Stratum of Large Units in Survey Design* = The American Statistician 1986. 1. sz. 27-31. p.
- [28.] **Hinkley, DV:** *On Power Transformations to Symmetry* = Biometrika 1975. Vol. 52. No. 1. 101-111. p.
- [29.] **Holt, D. – Smith, T. M. F.:** *Post stratification* = Journal of Royal Statistical Society A, 1979. Vol. 142. No. 1. 33-46. p.
- [30.] **Hosking, JRM:** *L-moments: analysis and estimation of distributions using linear combinations of order statistics* = Journal of the Royal Statistical Society, Series B 1990. Vol. 62. No. 1. 105-124. p.
- [31.] **Hunyadi, László:** *A two-phase sampling design* = Pure Mathematics and Applications. 1991. Ser. C. Vol. 2. No 1. 95-111. p.
- [32.] **Hunyadi, László – Rappai, Gábor:** *Gondolatok a statisztikáról* = Statisztikai Szemle. 1999. 77. évf. 1. sz. 5-15. p.
- [33.] **Hunyadi, László – Mundruczó, György – Vita, László:** *Statisztika.* Budapest: Aula Kiadó, 2000
- [34.] **Hunyadi, László:** *Statisztikai következtetéselmélet közgazdászoknak.* Budapest. Központi Statisztikai Hivatal, 2001a 483 p.
- [35.] **Hunyadi, László:** *A mintavétel alapjai.* Budapest. SZÁMALK Kiadó, 2001b 96 p.
- [36.] **Hunyadi, László – Vita László:** *Statisztika közgazdászoknak.* Budapest: Központi Statisztikai Hivatal, 2002. 770 p.

- [37.] **Hunyadi**, László: *Grafikus ábrázolás a statisztikában* = Statisztikai Szemle 80. évf. 2002. 1. sz. 22-52. p.
- [38.] **Kapitány**, Balázs: *Mintavételi módszerek ritka populációk esetén* = Statisztikai Szemle. 2010. 88. évf. 7-8. sz. 739-754. old.
- [39.] **Kerékgyártó**, Györgyné – **Mundruczó**, György: *Gazdasági szintű hatékonysági mutató képzése a mezőgazdaságban* = Statisztikai Szemle. 1976. 54. évf. 7. sz. 713-725. p.
- [40.] **Kish**, Leslie: *A mintavételek körüli százéves háború* = Statisztikai Szemle 73. évf. 1995a 10. sz. 773-787. p.
- [41.] **Kish**, Leslie: *Survey Sampling*. New York: John Wiley & Sons, INC. 1995b 643 p.
- [42.] **Kish**, Leslie: *Sampling methods for agricultural surveys*. FAO Statistical Development Series 3. Rome: Food and Agriculture Organization of the United Nations, 1989. 261 p.
- [43.] **Köves**, Pál – **Párniczky**, Gábor: *Általános statisztika I. Közgazdasági és Jogi Könyvkiadó, Budapest, 1981. 363 p.*
- [44.] **Köves**, Pál – **Párniczky**, Gábor: *Általános statisztika II. Közgazdasági és Jogi Könyvkiadó, Budapest, 1981. 387 p.*
- [45.] **Laczka**, Sándorné – **Oros**, Iván – **Schindele**, Miklós: *Magyarország állattenyésztése 1851-1996-ig (Statisztikai áttekintés)*, Központi Statisztikai Hivatal, Budapest, 1998. 101 p.
- [46.] **Laczka**, Sándorné: *Mezőgazdasági összeírások Magyarországon, 1895-2000* = Statisztikai Szemle 78. évf. 2000. 4. sz. 283-289. p.
- [47.] **Laczka**, Sándorné – **Szabó**, P.: *A gazdaság fogalma Magyarország és az EU mezőgazdasági statisztikájában* = Statisztikai Szemle 78. évf. 2000. 4. sz. 225-238. p.
- [48.] **Laczka**, Sándorné: *Ültetvény összeírások Magyarországon* = Statisztikai Szemle 80. évf. 2002. 7. sz. 648-662. p.



- [49.] **Mandrekar**, Jayawant N. – **Mandrekar**, Sumithra J. – **Cha**, Stephen S.: *Evaluating Methods of Symmetry*. Technical Report, Division of Biostatistics, Mayo Clinic, Rochester, 2005. 16 p.
- [50.] **Marton**, Ádám: *A reprezentatív felvételek megbízhatósága*. KSH Könyvtár és Dokumentációs Szolgálat, Budapest, 1991. 119 p.
- [51.] **McCallion**, T.: *Optimum Allocation in Stratified Random Sampling with Ratio Estimation as Applied to the Northern Ireland December Agricultural Sample* = Applied Statistics 1992. Vol. 41. No. 1. 39-45. p.
- [52.] **Mullahy, John**: *Heterogeneity, excess zeros, and the structure of count data models*. = Journal of Applied Econometrics. 1997. Vol. 12. Issue 3. 337-350. p.
- [53.] **Mundruczó**, György: *Sztochasztikus modellek közgazdasági alkalmazásának kérdései, különös tekintettel a mérési hibákra*. Kandidátusi értekezés. Budapest. 1979. 164 p.
- [54.] **Nyitrai**, Ferencné: *Gazdaságstatisztika*. Központi Statisztikai Hivatal, Budapest, 1996. 287 p.
- [55.] **Oros**, Iván: *A mezőgazdasági statisztika időszerű kérdései* = Statisztikai Szemle. 32. évf. 1954. 6-7. sz. 461-472. p.
- [56.] **Oros**, Iván: *Új tendenciák a mezőgazdasági kistermelésben* = Statisztikai Szemle 70. évf. 1992. 8. sz. 797-810. p.
- [57.] **Oros**, Iván: *Átfogó censusok a mezőgazdaságban* = Statisztikai Szemle 77. évf. 1999. 5. sz. 446-458. p.
- [58.] **Oros**, Iván: *A birtokszerkezet Magyarországon* = Statisztikai Szemle 80. évf. 2002. 7. sz. 674-697. p.
- [59.] **Párniczky**, Gábor: *A reprezentatív megfigyelés módszere és alkalmazásának egyes kérdései* = Statisztikai Szemle 32. évf. 1954. 4. sz. 271-283. p.
- [60.] **Párniczky**, Gábor – **Csepinszky**, Andor: *Reprezentatív megfigyelés a gazdasági statisztikában*. Közgazdasági és Jogi Könyvkiadó. Budapest. 1956. 255 p.

- [61.] **Párniczky, Gábor**: *A reprezentatív megfigyeléseknél fellépő torzítás problémái*. Kandidátusi Értekezés. 1966. 159 p.
- [62.] **Pintér, József – Rappai, Gábor** (szerk.): *Statisztika*. Pécs. Pécsi Tudományegyetem Közgazdaságtudományi Kar, 2007. 508 p.
- [63.] **Pintér, József – Rappai, Gábor**: *A mintavételi tervek készítésének néhány gyakorlati megfontolása* = Marketing & Menedzsment 35. évf. 2001. 4. sz. 4-10. p.
- [64.] **Randles, Ronald H. – Fligner, Michael A. – Policello, George E. II – Wolfe, Douglas A.**: *An Asymptotically Distribution-Free Tests for Symmetry Versus Asymmetry* = Journal of the American Statistical Association. 1980. Vol. 75. No. 369. 168-172. p.
- [65.] **Rappai, Gábor**: *Üzleti statisztika Excellel*. Központi Statisztikai Hivatal. Budapest, 2001. 231 p.
- [66.] **Ridout, M.S. – Demetrio, C.G.B. – Hinde, J.P.**: *Models for counts data with many zeros*. Proceedings of the XIXth International Biometric Conference, Cape Town, Invited Papers, 1998. 179-192. p.
- [67.] **Royston, Patrick**: *Which measures of skewness and kurtosis are best?* = Statistics in Medicine, Vol. 11. 1991. 333-343. p.
- [68.] **Schindele, Miklós**: *A reprezentatív állatszámmlálások módszere* = Statisztikai szemle 67. évf. 1989. 4. sz. 363-376. p.
- [69.] **Sethi, V. K.**: *A Note on Optimum Stratification of Population for Estimating Population Means* = Australian Journal of Statistics. 5. évf. 1963. 1. sz. 20-33 p.
- [70.] **Staub, Kevin E. – Winkelmann, Rainer**: *Consistent estimation of zero inflated count models*. University of Zurich. 2011 [www.econ.uzh.ch/faculty/staub/zeroinflated.pdf](http://www.econ.uzh.ch/faculty/staub/zeroinflated.pdf) (Letöltés dátuma: 2012. május 8.)
- [71.] **Székelyi, Mária – Barna, Ildikó**: *Túlélőkészlet az SPSS-hez*. Budapest: Typotex Kiadó, 2002. 453 p.

- [72.] **Szemán**, Zsuzsanna – **Szép**, Katalin – **Tóth**, Péter: *Az állatállomány megfigyelés statisztikája (módszertani alapidokumentáció)*. Központi Statisztikai Hivatal, Budapest, 1998. 74 p.
- [73.] **Szép**, Katalin – **Vigh**, Judit: *A minőség a hivatalos statisztikában* = Statisztikai Szemle 82. évf. 2004. 8. sz. 773-798. p.
- [74.] **Szilágyi**, Roland: *Mintavételen alapuló kvantitatív következtetések hibáinak kezelése*. PhD értekezés. Miskolci Egyetem Gazdaságtudományi Kar. 2011. 198. p.
- [75.] *Agrárcenzusok – 2010. évi Általános Mezőgazdasági Összeírás (ÁMÖ 2010) – Nemzetközi kitekintés* ([www.ksh.hu/agrarcenzusok\\_amo\\_2010\\_nemzetkozi](http://www.ksh.hu/agrarcenzusok_amo_2010_nemzetkozi)) Letöltés dátuma: 2012. április 15.
- [76.] *Szőlőültetvények összeírása, 2009* = Statisztikai tükör. IV. évf. 2010. 73. sz. 2 p. ([www.ksh.hu](http://www.ksh.hu)) Letöltés dátuma: 2011. május 7.
- [77.] *Magyarország mezőgazdasága, 2010* (Általános mezőgazdasági összeírás) Előzetes adatok, Központi Statisztikai Hivatal, 2010. december. 17 p. (Internetes kiadvány - [www.ksh.hu](http://www.ksh.hu)) Letöltés dátuma: 2011. február 20.
- [78.] *Állatállomány, 2010. december 1.* = Statisztikai tükör V. évf. 2011. 3. sz. 4 p. ([www.ksh.hu](http://www.ksh.hu)) Letöltés dátuma: 2011. március 18.

## Felhasznált jogszabályok

- [1.] 2010. évi XXIV. törvény a 2010. évi általános mezőgazdasági összeírásról
- [2.] Az Európai Parlament és a Tanács 1166/2008/EK rendelete (2008. november 19.) a gazdaságszerkezeti felmérésekről és a mezőgazdasági termelési módszereket vizsgáló felmérésről, valamint az 571/88/EGK tanácsi rendelet hatályon kívül helyezéséről
- [3.] Az Európai Parlament és a Tanács 1165/2008/EK rendelete (2008. november 19.) az állatállományra és a húsokra vonatkozó statisztikákról, valamint a 93/23/EGK, 93/24/EGK és 93/25/EGK tanácsi irányelvek hatályon kívül helyezéséről

- [4.] A Tanács 571/88/EGK Rendelete (1988. február 29.) a mezőgazdasági üzemek szerkezetére vonatkozó, 1988. és 1997. évek közötti közösségi felmérések szervezéséről
- [5.] 1874. évi XXV. törvény-cikk az országos statistika ügyének szervezéséről
- [6.] 1897. évi XXXV. törvény-cikk a M. Kir. Központi Statisztikai hivatalról
- [7.] 1929. évi XIX. törvénycikk a hivatalos statisztikai szolgálatról
- [8.] 1952. évi VI. törvény az állami statisztikáról
- [9.] 1957. évi IV. törvény az államigazgatási eljárás általános szabályairól
- [10.] 1973.évi V. törvény a statisztikáról
- [11.] A Tanács 1990. június 11-i 1588/90/EURATOM, EGK rendelete a bizalmas statisztikai adatoknak az Európai Közösségek Statisztikai Hivatala részére történő átadásáról
- [12.] 78/1990. (IV.25.) MT rendelet a statisztikáról szóló 1973. évi V. törvény végrehajtásáról
- [13.] 1992. évi LXIII. törvény a személyes adatok védelméről és a közérdekű adatok nyilvánosságáról
- [14.] 1993. évi XLVI. törvény a statisztikáról, egységes szerkezetben a végrehajtásáról szóló 170/1993. (XII. 3.) Korm. rendelettel
- [15.] Az Európai Parlament és a Tanács 1995. október 24-i 95/46/EK irányelve a személyes adatok kezelése vonatkozásában az egyének védelméről és az ilyen adatok szabad áramlásáról
- [16.] A Tanács 1997. február 17-i 322/97/EK rendelete a közösségi statisztikákról
- [17.] A Bizottság 1997. április 21-i 97/281/EK határozata az Eurostat szerepéről a közösségi statisztikák elkészítésében
- [18.] 1998. évi VI. törvény az egyének védelméről a személyes adatok gépi feldolgozása során, Strasbourgban, 1981. január 28. napján kelt Egyezmény kihirdetéséről

## I. számú melléklet

### Állatállomány alakulása, 1895-2010 (ezer db)

Év	Referencia időpont	Szarvasmarha	Sertés	Juh
1895. (ÁMÖ)	november 20.	2 062	3 180	3 151
1911.	február 28.	2 185	3 352	2 446
1922.	május 31.	1 820	2 473	1 352
1923.	május 31.	1 810	2 133	1 587
1924.	május 31.	1 887	2 458	1 814
1925.	május 31.	1 911	2 631	1 890
1926.	május 31.	1 838	2 518	1 803
1927.	május 31.	1 796	2 384	1 610
1928.	május 31.	1 802	2 659	1 566
1929.	május 31.	1 810	2 580	1 573
1930.	május 1.	1 776	2 362	1 463
1931.	május 31.	1 805	2 712	1 440
1932.	május 31.	1 810	2 359	1 210
1933.	május 31.	1 688	1 898	1 056
1934.	május 31.	1 669	2 500	1 087
1935. (ÁMÖ)	február 28.	1 911	4 674	1 450
1936.	május 31.	1 732	2 552	1 350
1937.	május 31.	1 747	2 621	1 483
1938.	március 15.	1 872	3 107	1 628
1942. (ÁMÖ)	február 28.	2 363	4 667	1 708
1945.	november 30.	1 059	1 113	328
1946.	május 31.	1 100	1 327	370
1947.	május 31.	1 841	2 792	570
1948.	május 31.	1 993	2 771	579
1949.	február 15.	1 942	3 316	910
1950.	február 20.	2 222	5 542	1 049
1951.	február 28.	2 009	4 298	1 143
1952.	március 3.	2 091	4 740	1 481
1953.	március 2.	2 236	4 977	1 637
1954.	március 1.	2 076	4 454	1 869
1955.	március 1.	2 130	5 817	1 858
1956.	március 1.	2 169	6 055	1 928
1957.	március 1.	1 973	4 996	1 873
1958.	március 1.	1 937	5 338	2 049
1959.	március 1.	2 004	6 225	2 155
1960. (ÁMÖ)	március 1.	1 971	5 356	2 381
1961.	március 7.	1 957	5 921	2 643
1962.	március 15.	1 987	6 409	2 850
1963.	március 31.	1 906	5 428	3 043
1964.	március 31.	1 883	6 358	3 305
1965.	március 31.	1 964	6 963	3 400
1966.	március 31.	1 973	5 799	3 270
1967.	március 31.	2 014	6 005	3 274
1968.	március 31.	2 096	6 609	3 311
1969.	március 31.	2 006	5 334	3 277

1970. (ÁMÖ)	március 31.	1 933	5 970	3 024
1971.	március 31.	1 917	7 510	2 657
1972.	március 31.	1 901	7 353	2 271
1973.	március 31.	1 965	6 980	2 259
1974.	március 31.	2 035	9 368	2 021
1975.	március 31.	2 041	7 885	2 039
1976.	március 31.	1 941	7 175	2 350
1977.	március 31.	1 960	8 282	2 619
1978.	március 31.	2 020	8 024	2 863
1979.	március 31.	2 005	8 423	2 927
1980. (ÁMÖ)	március 31.	1 960	8 570	3 090
1981.	március 31.	1 976	8 336	3 140
1982.	március 31.	2 007	8 757	3 180
1983.	március 31.	1 977	9 799	2 977
1984.	március 31.	1 982	9 858	2 832
1985.	március 31.	1 948	9 168	2 465
1986.	március 31.	1 824	8 649	2 337
1987.	március 31.	1 768	8 936	2 336
1988.	március 31.	1 715	8 664	2 216
1989.	március 31.	1 730	8 553	2 069
1990.	március 31.	1 637	8 457	1 865
1991. (ÁMÖ)	március 31.	1 592	8 590	2 289
1992.	március 31.	1 369	6 175	2 076
1993.	március 31.	1 114	5 813	1 598
1994.	március 31.	984	5 035	1 237
1995.	március 31.	951	4 669	1 074
1996.	december 1.	909	5 289	872
1997.	december 1.	871	4 931	858
1998.	december 1.	873	5 479	909
1999.	december 1.	857	5 335	934
2000. (ÁMÖ)	december 1.	805	4 834	1 129
2001.	december 1.	783	4 822	1 136
2002.	december 1.	770	5 082	1 103
2003.	december 1.	739	4 913	1 296
2004.	december 1.	723	4 059	1 397
2005.	december 1.	708	3 853	1 405
2006.	december 1.	702	3 987	1 298
2007.	december 1.	705	3 871	1 232
2008.	december 1.	701	3 383	1 236
2009.	december 1.	700	3 247	1 223
2010. (ÁMÖ)	december 1.	682	3169	1181

Forrás: KSH honlap „Agrár idősorok és cenzusok” (Hosszú idősorok)

## II. számú melléklet

### Magyarország földterületének alakulása néhány művelési ágban, 1950-2010 (ezer hektár)

Év	Szántó	Gyümölcsös	Szőlő	Gyep	Erdő
1950	5 518,10	58,1	230,2	1 474,70	1 165,90
1951	5 518,70	57,6	228,5	1 455,30	1 165,30
1952	5 527,30	59,1	225,3	1 441,80	1 172,60
1953	5 492,80	58,9	215,8	1 421,40	1 248,70
1954	5 466,90	62,4	215,8	1 441,50	1 248,80
1955	5 402,90	64,9	201,4	1 471,40	1 257,40
1956	5 395,20	66,3	195,6	1 450,20	1 263,10
1957	5 388,70	66,7	195,7	1 440,90	1 271,60
1958	5 378,10	69,4	198,5	1 440,80	1 272,50
1959	5 369,60	70,2	200,7	1 434,90	1 277,50
1960 (ÁMÖ)	5 309,80	82,3	203,6	1 437,90	1 306,20
1961	5 208,00	97,8	204,4	1 459,30	1 334,20
1962	5 135,20	124,2	219	1 390,30	1 367,60
1963	5 106,40	134,9	229,3	1 359,30	1 389,20
1964	5 086,20	153,4	242,6	1 341,50	1 400,90
1965	5 084,50	167,9	246,6	1 303,90	1 421,50
1966	5 076,70	172,4	244,9	1 285,30	1 441,90
1967	5 065,60	171,6	239,7	1 287,80	1 450,80
1968	5 058,00	170	236,1	1 289,90	1 455,20
1969	5 053,60	170,7	233,5	1 283,60	1 463,00
1970 (ÁMÖ)	5 046,20	171,6	229,7	1 281,30	1 470,70
1971	5 033,00	171,8	222,3	1 276,80	1 479,80
1972	5 026,10	169,5	218,5	1 281,20	1 483,20
1973	5 025,00	165,3	213,3	1 280,10	1 489,50
1974	4 977,70	163,4	210,4	1 279,20	1 536,00
1975	4 975,60	161,1	206,2	1 274,80	1 545,30
1976	4 953,70	163	199,8	1 286,50	1 555,60
1977	4 908,80	159,9	191,8	1 306,80	1 574,60
1978	4 868,50	156,6	185,7	1 309,30	1 584,30
1979	4 795,60	143,5	173,8	1 296,60	1 594,00
1980 (ÁMÖ)	4 734,70	138,4	167,8	1 294,20	1 610,30
1981	4 687,80	125	161,3	1 284,00	1 620,20
1982	4 679,30	121,2	159,2	1 283,30	1 626,70
1983	4 681,30	114,8	156,7	1 279,20	1 632,70
1984	4 687,90	107,9	154,7	1 264,90	1 639,70
1985	4 697,50	103,5	153,6	1 246,40	1 647,90
1986	4 704,90	99	147,4	1 233,70	1 659,20
1987	4 709,30	96,5	144,9	1 222,30	1 669,00
1988	4 712,00	94,9	142,2	1 209,90	1 678,30
1989	4 712,70	94,3	140,3	1 197,30	1 688,20
1990	4 712,80	95,1	138,5	1 185,60	1 695,40
1991 (ÁMÖ)	4 714,20	94,1	136,5	1 173,00	1 701,20
1992	4 706,90	94,5	135	1 164,00	1 712,20
1993	4 712,50	93	131,7	1 156,60	1 763,90

1994	4 714,40	92,7	131,9	1 148,00	1 766,50
1995	4 715,90	93,9	131,3	1 148,00	1 762,90
1996	4 712,70	94,3	130,9	1 148,40	1 764,50
1997	4 710,80	95,6	130,9	1 148,10	1 766,70
1998	4 709,50	96,3	129,7	1 147,80	1 769,30
1999	4 708,00	96,4	127	1 147,20	1 774,90
2000 (ÄMÖ)	4 499,80	95,4	105,9	1 051,20	1 769,60
2001	4 516,10	97,5	92,9	1 061,20	1 771,70
2002	4 515,50	97,3	92,8	1 063,10	1 771,70
2003	4 515,50	98,3	93,3	1 061,60	1 775,10
2004	4 510,30	102,6	94,5	1 059,60	1 775,10
2005	4 513,10	102,8	86	1 056,90	1 775,10
2006	4 509,60	102,8	86	1 014,50	1 776,70
2007	4 506,10	101,9	86	1 016,90	1 822,40
2008	4 502,80	98,5	82,6	1 009,80	1 884,40
2009	4 501,60	98,7	82,8	1 004,20	1 895,60
2010 (ÄMÖ)	4 501,60	93,7	82,8	762,6	1 912,90



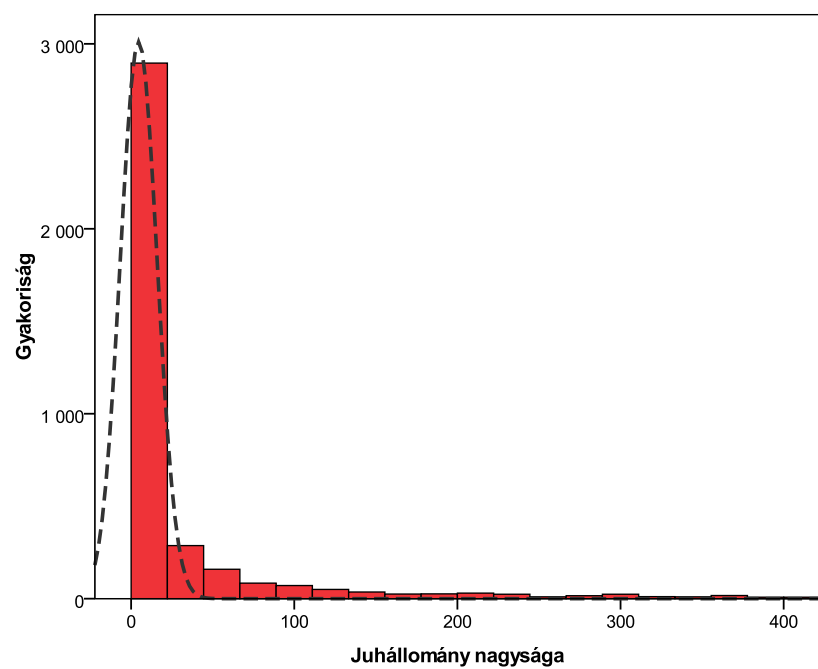
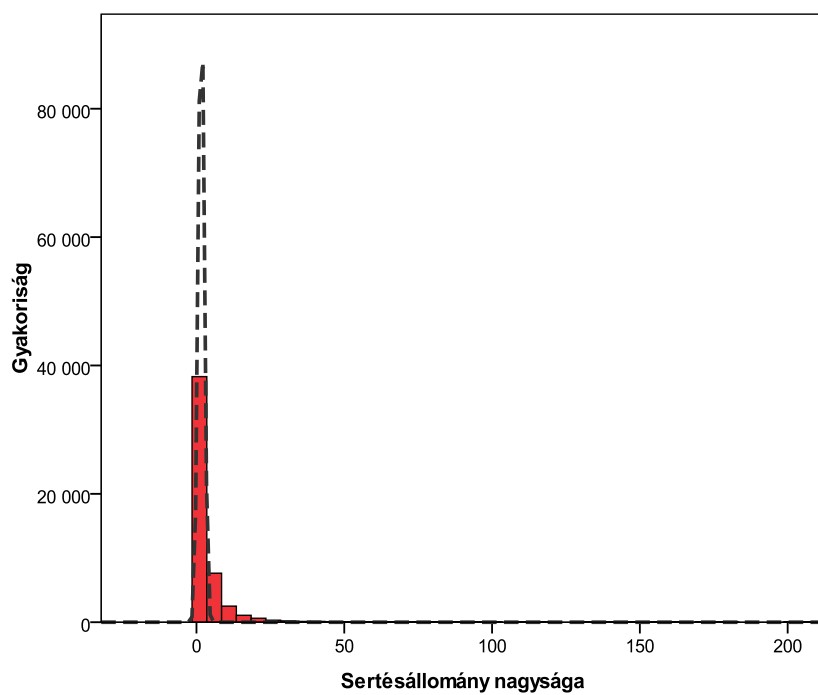
### III. számú melléklet

#### Mezőgazdasági témájú KSH-s adatgyűjtések

OSAP	A kérdőív címe	Gyakoriság
1082	Földterület és vetésterület	évközi
1084	A kalászos gabonák terméseredménye	évközi
1085	Főbb növénykultúrák terméseredménye	évközi
1087	Állatállomány	évközi (2-szer)
1089	Egyéni gazdaságok állatállománya	évközi (2-szer)
1097	Jelentés a felvásárlásról	havi
1099	Áruforgalmi mérleg	havi
1107	Állatvásári jelentés	havi
1108	Piaci felhozatal és árjelentés	havi
1651	Az egyéni gazdaságok földterületének és vetésterületének összeírása	évközi
1677	Egyéni gazdaságok mezőgazdasági összeírása	éves
1758	A mezőgazdasági és erdőgazdálkodási termelők ráfordításai	havi
1791	Baromfikeltetés	havi
1824	Takarmányok értékesítési ára	havi
1826	Növényvédő szerek értékesítési ára	havi
1827	Állatgyógyászati készítmények értékesítési ára	havi
1869	Tehéntej felvásárlása és tejtermékek előállítása	havi
1870	Tej- és tejtermékek termelési, felhasználása és nettó árbevétele	éves
1871	Mezőgazdasági és erdőgazdálkodási szolgáltatások	éves
2128	Gazdasági szervezetek éves mezőgazdasági adatai	éves

## IV. számú melléklet

### A sertések és juhok hisztogramja



## V. számú melléklet

**Az egyéni gazdaságok számának alakulása a tipológia csoportjai és régiók szerint, 2007**

Tipológia\Régió	Közép-Magyarország	Közép-Dunántúl	Nyugat-Dunántúl	Dél-Dunántúl	Észak-Magyarország	Észak-Alföld	Dél-Alföld	Összesen
Szántóföldi növénytermesztés	7 725	7 229	6 034	10 076	7 497	31 423	23 072	<b>93 055</b>
Kertészet	749	154	318	157	737	1 133	4 626	<b>7 874</b>
Állandó növénykultúrák termesztése	7 091	11 810	9 966	10 210	19 609	20 045	13 325	<b>92 057</b>
Legeltetési állattartás	1 952	2 177	2 073	1 998	3 255	5 269	5 293	<b>22 016</b>
Abrakfogyasztó állatok tartása	19 098	10 709	11 132	18 696	14 159	28 705	36 943	<b>139 441</b>
Vegyes növénytermesztés	4 718	5 116	9 384	7 832	8 290	18 907	11 130	<b>65 377</b>
Vegyes állattartás	10 533	8 883	10 254	13 389	13 566	18 746	18 107	<b>93 477</b>
Növénytermesztés és állattartás vegyesen	7 589	8 205	11 181	12 877	13 354	22 807	18 599	<b>94 613</b>
Nem besorolható	1 813	891	985	1 632	528	1 936	2 971	<b>10 756</b>
<b>Összesen</b>	<b>61 267</b>	<b>55 173</b>	<b>61 326</b>	<b>76 867</b>	<b>80 995</b>	<b>148 971</b>	<b>134 066</b>	<b>618 665</b>

## VI. számú melléklet

### Az egyes állatfajok jellemzői régiók szerint

Régió	Szórás (állatszám)			Átlag (állatszám)			Relatív szórás (%)		
	Szarvas-marha	Sertés	Juh	Szarvas-marha	Sertés	Juh	Szarvas-marha	Sertés	Juh
Közép-Magyarország	2,01	4,28	16,61	0,15	1,27	0,69	1301	338	2408
Közép-Dunántúl	2,40	9,86	15,26	0,23	1,83	0,79	1058	538	1928
Nyugat-Dunántúl	2,79	11,55	14,81	0,30	2,02	0,11	927	573	12942
Dél-Dunántúl	2,46	7,52	2,48	0,21	1,87	0,73	1199	402	340
Észak-Magyarország	1,63	3,84	14,70	0,13	1,01	0,79	1239	382	1856
Észak-Alföld	2,34	6,52	24,35	0,24	1,79	2,22	954	365	1099
Dél-Alföld	3,02	11,66	21,18	0,38	3,31	1,77	803	352	1195

## VII. számú melléklet

### Belső szórás különböző csoportosítások mellett

Állatfaj	Réteghatár (db)	Belső szórás	
		Állatállományt tartók	Összes gazdaság
Sertés	10	8,31	7,08
	30	7,48	6,42
	50	7,17	6,18
	70	7,03	6,06
	80	6,99	6,04
	100	6,92	5,98
	120	6,93	5,99
	140	6,98	6,03
Szarvasmarha	5	1,96	1,67
	10	1,71	1,45
	15	1,64	1,40
	20	1,67	1,42
	30	1,83	1,56
	50	2,23	1,90
Juh	20	16,73	14,19
	50	14,25	12,09
	90	12,62	10,71
	120	11,88	10,08
	140	11,60	9,85
	180	11,47	9,73
	200	11,58	9,83
	220	11,71	9,94

## VIII. számú melléklet

Gyakorisági sorok a kumulatív  $\sqrt{f}$  szabály alkalmazásához

### Szarvasmarha

Osztályköz (db)	Gyakoriság (f)	$\sqrt{f}$	Kumulált $\sqrt{f}$
0-5	78 483	280,1	280,1
6-10	656	25,6	305,8
11-15	222	14,9	320,7
16-20	136	11,7	332,3
21-25	93	9,6	342,0
26-30	59	7,7	349,6
31-35	43	6,6	356,2
36-40	28	5,3	361,5
41-45	15	3,9	365,4
46-50	21	4,6	370,0
51-55	20	4,5	374,4
56-60	13	3,6	378,0
61-65	10	3,2	381,2
66-70	9	3,0	384,2
71-75	2	1,4	385,6
76-80	8	2,8	388,4
81-85	4	2,0	390,4
86-90	1	1,0	391,4
91-95	2	1,4	392,8

### Sertés

Osztályköz (db)	Gyakoriság (f)	$\sqrt{f}$	Kumulált $\sqrt{f}$
0-10	75721	275,2	275,2
11-20	2599	51,0	326,2
21-30	737	27,1	353,3
31-40	294	17,1	370,4
41-50	135	11,6	382,1
51-60	87	9,3	391,4
61-70	63	7,9	399,3
71-80	40	6,3	405,7
81-90	27	5,2	410,9
91-100	28	5,3	416,1
101-110	7	2,6	418,8
111-120	8	2,8	421,6
121-130	4	2,0	423,6
131-140	9	3,0	426,6
141-150	7	2,6	429,3
151-160	10	3,2	432,4
161-170	3	1,7	434,2
171-180	5	2,2	436,4
181-190	0	0,0	436,4
191-200	2	1,4	437,8

201-210	3	1,7	439,5
211-220	3	1,7	441,3
221-230	1	1,0	442,3
231-240	4	2,0	444,3
241-250	2	1,4	445,7
251-260	2	1,4	447,1
261-270	1	1,0	448,1
271-280	1	1,0	449,1
281-290	2	1,4	450,5
291-300	0	0,0	450,5
301-310	1	1,0	451,5
311-320	2	1,4	452,9
321-330	1	1,0	453,9
331-340	0	0,0	453,9
341-350	3	1,7	455,7
351-360	1	1,0	456,7
361-370	1	1,0	457,7
371-380	0	0,0	457,7
381-390	0	0,0	457,7
391-400	1	1,0	458,7
401-410	4	2,0	460,7
411-420	1	1,0	461,7
421-430	2	1,4	463,1
431-440	0	0,0	463,1
441-450	1	1,0	464,1
451-460	1	1,0	465,1
461-470	0	0,0	465,1
471-480	1	1,0	466,1

### Juh

Osztályköz (db)	Gyakoriság (f)	$\sqrt{f}$	Kumulált $\sqrt{f}$
0-10	78335	279,9	279,9
11-20	490	22,1	302,0
21-30	189	13,7	315,8
31-40	125	11,2	326,9
41-50	97	9,8	336,8
51-60	65	8,1	344,9
61-70	35	5,9	350,8
71-80	42	6,5	357,3
81-90	35	5,9	363,2
91-100	36	6,0	369,2
101-110	23	4,8	374,0
111-120	26	5,1	379,1
121-130	21	4,6	383,6
131-140	18	4,2	387,9
141-150	17	4,1	392,0
151-160	12	3,5	395,5
161-170	14	3,7	399,2
171-180	12	3,5	402,7
181-190	15	3,9	406,6
191-200	15	3,9	410,4

201-210	13	3,6	414,0
211-220	6	2,4	416,5
221-230	14	3,7	420,2
231-240	8	2,8	423,1
241-250	6	2,4	425,5
251-260	3	1,7	427,2
261-270	6	2,4	429,7
271-280	6	2,4	432,1
281-290	7	2,6	434,8
291-300	9	3,0	437,8
301-310	15	3,9	441,7
311-320	4	2,0	443,7
321-330	7	2,6	446,3
331-340	2	1,4	447,7
341-350	7	2,6	450,4
351-360	9	3,0	453,4
361-370	6	2,4	455,8
371-380	4	2,0	457,8
381-390	5	2,2	460,0
391-400	8	2,8	462,9
401-410	0	0,0	462,9
411-420	2	1,4	464,3
421-430	1	1,0	465,3
431-440	3	1,7	467,0
441-450	3	1,7	468,8
451-460	1	1,0	469,8
461-470	1	1,0	470,8
471-480	4	2,0	472,8
491-500	4	2,0	476,2
501-510	3	1,7	477,9
511-520	3	1,7	479,6
521-530	0	0,0	479,6
531-540	1	1,0	480,6
541-550	1	1,0	481,6
551-560	3	1,7	483,4
561-570	5	2,2	485,6
571-580	0	0,0	485,6
581-590	0	0,0	485,6
591-600	2	1,4	487,0
601-610	0	0,0	487,0
611-620	0	0,0	487,0
621-630	0	0,0	487,0
631-640	0	0,0	487,0
641-650	1	1,0	488,0
651-660	0	0,0	488,0
661-670	1	1,0	489,0
671-680	3	1,7	490,7
681-690	1	1,0	491,7
691-700	2	1,4	493,2
701-710	2	1,4	494,6
711-720	0	0,0	494,6
721-730	1	1,0	495,6
731-740	1	1,0	496,6



741-750	0	0,0	496,6
751-760	2	1,4	498,0
761-770	1	1,0	499,0
771-780	0	0,0	499,0
781-790	0	0,0	499,0
791-800	0	0,0	499,0
801-810	0	0,0	499,0
811-820	1	1,0	500,0
821-830	0	0,0	500,0
831-840	0	0,0	500,0
841-850	1	1,0	501,0
851-860	0	0,0	501,0
861-870	0	0,0	501,0
871-880	0	0,0	501,0
881-890	1	1,0	502,0
891-900	0	0,0	502,0
901-910	0	0,0	502,0
911-920	0	0,0	502,0
921-930	0	0,0	502,0
931-940	0	0,0	502,0
941-950	0	0,0	502,0
951-960	2	1,4	503,4

## IX. számú melléklet

### Az exponenciális illesztéshez felhasznált adatok

Szarvas- marha szám	Gyakoriság	Értékösszeg	Relatív gyakoriság	Exponenciális eloszlás értéke	(relatív gyakoriság - exponenciális eloszlás értéke) <sup>2</sup> /exponenciális eloszlás értéke
1	538	538	0,1561	0,1087	0,0206
2	591	1182	0,1715	0,0962	0,0590
3	417	1251	0,1210	0,0850	0,0152
4	348	1392	0,1010	0,0752	0,0088
5	211	1055	0,0612	0,0665	0,0004
6	212	1272	0,0615	0,0588	0,0001
7	142	994	0,0412	0,0520	0,0022
8	114	912	0,0331	0,0460	0,0036
9	93	837	0,0270	0,0407	0,0046
10	95	950	0,0276	0,0360	0,0020
11	46	506	0,0133	0,0318	0,0107
12	57	684	0,0165	0,0281	0,0048
13	38	494	0,0110	0,0249	0,0077
14	41	574	0,0119	0,0220	0,0046
15	40	600	0,0116	0,0194	0,0032
16	27	432	0,0078	0,0172	0,0051
17	29	493	0,0084	0,0152	0,0030
18	29	522	0,0084	0,0134	0,0019
19	20	380	0,0058	0,0119	0,0031
20	31	620	0,0090	0,0105	0,0002
21	19	399	0,0055	0,0093	0,0015
22	23	506	0,0067	0,0082	0,0003
23	12	276	0,0035	0,0073	0,0020
24	21	504	0,0061	0,0064	0,0000
25	18	450	0,0052	0,0057	0,0000
26	15	390	0,0044	0,0050	0,0001
27	7	189	0,0020	0,0044	0,0013
28	16	448	0,0046	0,0039	0,0001
29	10	290	0,0029	0,0035	0,0001
30	11	330	0,0032	0,0031	0,0000
31	9	279	0,0026	0,0027	0,0000
32	12	384	0,0035	0,0024	0,0005
33	6	198	0,0017	0,0021	0,0001
34	7	238	0,0020	0,0019	0,0000
35	9	315	0,0026	0,0017	0,0005
36	7	252	0,0020	0,0015	0,0002
37	8	296	0,0023	0,0013	0,0008
38	4	152	0,0012	0,0011	0,0000
39	5	195	0,0015	0,0010	0,0002
40	4	160	0,0012	0,0009	0,0001
41	2	82	0,0006	0,0008	0,0001

42	2	84	0,0006	0,0007	0,0000
43	3	129	0,0009	0,0006	0,0001
44	1	44	0,0003	0,0005	0,0001
45	7	315	0,0020	0,0005	0,0049
46	4	184	0,0012	0,0004	0,0012
47	4	188	0,0012	0,0004	0,0016
48	5	240	0,0015	0,0003	0,0037
49	2	98	0,0006	0,0003	0,0003
50	6	300	0,0017	0,0003	0,0083
51	5	255	0,0015	0,0002	0,0064
52	3	156	0,0009	0,0002	0,0022
53	2	106	0,0006	0,0002	0,0009
54	6	324	0,0017	0,0002	0,0155
55	4	220	0,0012	0,0001	0,0073
56	3	168	0,0009	0,0001	0,0044
57	1	57	0,0003	0,0001	0,0003
58	4	232	0,0012	0,0001	0,0115
59	2	118	0,0006	0,0001	0,0028
60	3	180	0,0009	0,0001	0,0082
61	3	183	0,0009	0,0001	0,0095
62	3	186	0,0009	0,0001	0,0109
65	4	260	0,0012	0,0000	0,0301
66	3	198	0,0009	0,0000	0,0189
68	3	204	0,0009	0,0000	0,0247
69	2	138	0,0006	0,0000	0,0121
70	1	70	0,0003	0,0000	0,0032
72	1	72	0,0003	0,0000	0,0042
75	1	75	0,0003	0,0000	0,0064
76	1	76	0,0003	0,0000	0,0073
77	3	231	0,0009	0,0000	0,0780
79	1	79	0,0003	0,0000	0,0108
80	3	240	0,0009	0,0000	0,1136
81	2	162	0,0006	0,0000	0,0568
82	1	82	0,0003	0,0000	0,0158
85	1	85	0,0003	0,0000	0,0231
88	1	88	0,0003	0,0000	0,0337
91	1	91	0,0003	0,0000	0,0490
93	1	93	0,0003	0,0000	0,0628

## Publikációs jegyzék

### Cikkek folyóiratokban, konferenciakötetekben:

1. **Galambosné Tiszberger, Mónika:** *Magyarország szerepe a LUCAS-felvételek módszertanának kidolgozásában.* In: Herman, Sándor (szerk.): Magyar Statisztikai Társaság konferenciája, Balatonfüred, 2005. október 20-21. 63-72. p.
2. **Galambosné Tiszberger, Mónika:** *A Földhasználati és földfelszín-borítottsági összeírás módszertani háttere.* = Statisztikai Szemle 2006. 9. szám 829-847. p.
3. **Galambosné Tiszberger, Mónika:** *Sampling techniques in agricultural statistics.* In: Dohnal, G. – Antoch, J. (szerk.): Students' conference of Current Trends in Statistics in V6 region, Prága, 2008. szeptember 4-5. 36-44. p.
4. **Galambosné Tiszberger, Mónika:** *Kisgazdaságok a magyar mezőgazdaságban.* In: Buday-Sántha, Attila (szerk.): Évkönyv 2009, Pécsi Tudományegyetem Közgazdaságtudományi Kar, Regionális Politika és Gazdaságtan Doktori Iskola, 2009. 60-69. p.
5. **Galambosné Tiszberger, Mónika:** *Sampling techniques for sampling units with different size.* In: Kovács, Péter – Szép, Katalin – Katona, Tamás (szerk.): Social and Business processes, International conference, Szeged, 2009. november 19-21. 351-370. p.
6. **Galambosné Tiszberger, Mónika:** *Megyei és regionális szintű adatok a mezőgazdaság statisztikában (A területi adatok módszertani háttere).* In: Buday-Sántha, Attila – Erdősi, Ferenc – Horváth, Gyula: Évkönyv 2010 II. kötet. „Félidőben” A közép-európai terület-, település-, vidék- és környezetfejlesztéssel foglalkozó doktori iskolák találkozik és konferenciája, IV. Országos környezetgazdaságtani PhD-konferencia, Pécsi Tudományegyetem Közgazdaságtudományi Kar, Regionális Politika és Gazdaságtan Doktori Iskola, Pécs, 2010. 164-174. p.
7. **Leidecker, Eleonóra – Galambosné Tiszberger, Mónika – Kráncz, János:** *Gyalogló és ülő foglalkozású populáció vizsgálata, a fizikai aktivitás kapcsolata a mozgatórendszer panaszaival.* Magyar Epidemiológia, 2011. VIII. évf. 9. sz. 13-20. p.
8. **Kuráth, Gabriella – Kovács, Árpád – Kiss, Tibor – Szűcs, Krisztián – Tiszberger, Mónika – Héráné Tóth, Andrea – Sipos, Norbert:** *Diplomás Pályakövető Rendszer – 2010-es pályakövetési vizsgálat a Pécsi Tudományegyetemen.* Pécsi Tudományegyetem. Pécs, 2011. február ISSN: 2062-5618 48 o. (www.pte.hu/DPR) (Készült a Pécsi Tudományegyetem RH Marketing Osztály gondozásában.)
9. **Kuráth, Gabriella – Kovács, Árpád – Kiss, Tibor – Szűcs, Krisztián – Tiszberger, Mónika – Héráné Tóth, Andrea – Sipos, Norbert:** *Diplomás Pályakövető Rendszer –*

*2010-es motivációs vizsgálat a Pécsi Tudományegyetemen.* Pécsi Tudományegyetem. Pécs, 2011. február ISSN: 2062-5618 57 o. (www.pte.hu/DPR)(Készült a Pécsi Tudományegyetem RH Marketing Osztály gondozásában.)

10. **Galambosné Tiszberger, Mónika:** *Rétegzési lehetőségek az állatállomány összeírásoknál.* In: Lázár, Ede (szerk.): Gazdasági és üzleti kihívások a Kárpát-medencében című konferencia kiadványkötete. Sapientia Egyetem. Státus Kiadó, Csíkszereda, 2011. 132-142. p.
11. **Galambosné Tiszberger, Mónika:** *A rétegzett mintavételről.* = Statisztikai Szemle 2011. 89. évf. 9. sz. 909-929. p.
12. **Kerekes, Zsuzsanna, Tiszberger Mónika, Kiss Edina, Bors P., Tóth M., Gáti Ágnes, Mezősi Emese:** *A személyiség, pszichológiai immunitás és depresszió szerepe krónikus betegeknél* (IBD, autoimmun pajzsmirigy betegség és psoriasis). In: Gy. Kiss E., Polyák L. (szerk.): Fogyatékossgal élők közösségfejlesztő gyakorlata és módszerei Közép-Európában. Pécs. A Magyar ILCO Szövetség 2011-es konferenciájának válogatott, lektorált tanulmányai. Kaposvár, Magyar ILCO Szövetség, 2011. 83-94. p.
13. **Szűcs, Krisztián – Galambosné Tiszberger, Mónika – Kuráth, Gabriella:** *Ilyenek vagytok Ti - egy korrespondenciaelemzés tanulságai.* = Marketing & Menedzsment 2011. (megjelenés alatt)

#### **Konferencia előadások:**

1. **Tiszberger, Mónika:** *LUCAS implementation in Hungary.* LUCAS-Phare Workshop, Eurostat, Luxemburg, 2003. január 20.
2. **Tiszberger, Mónika:** *LUCAS survey in Hungary, 2002.* Working Group (Land use statistics) meeting, Eurostat, Luxemburg, 2003. március 10-11.
3. **Leidecker, Eleonóra – Galambosné Tiszberger, Mónika – Kellerman, Péter – Kráncz, János:** *Fizikai aktivitás és inaktivitás hatása az ízületekre, gyalogló védőnők vizsgálata,* XXI. Országos Szülésznő-Védőnő-Gyermekápoló Konferencia, Budapest, 2010. november 23-24.
4. **Galambosné Tiszberger, Mónika – Szűcs, Krisztián:** *Karok hallgatói megítélése a Pécsi Tudományegyetemen,* III. SPSS Nyári Iskola, Szeged, 2011. július 11-14.

## Poszterek:

1. **Kiss, Edina – Kerekes, Zsuzsanna – Mezősi, Emese – Tiszberger, Mónika – Gáti, Ágnes:** *Az autoimmun pajzsmirigy betegségek pszichés vonatkozásai. A stressztől a boldogságig.* XI. Magatartástudományi Napok, Gödöllő, 2011. június 28-29.
2. **Kopasz, Nóra – Kerekes, Zsuzsanna – Mangel, László – Édes, Tünde – Tiszberger, Mónika – Pálfiné Szabó, Ilona:** *Kiegészítő tünetek és diszfunkcionális attitűdök vizsgálata egészségügyi dolgozóknál. A stressztől a boldogságig.* XI. Magatartástudományi Napok, Gödöllő, 2011. június 28-29.
3. **Leidecker, Eleonóra – Galambosné Tiszberger, Mónika – Bohner-Beke, Alíz – Kráncz, János:** *Connection between physical activity and complaints of musculoskeletal system.* 7th EFSMA-European Sports Medicine Congress, 3rd Central European Congress of Physical Medicine and Rehabilitation, Salzburg, 2011. október 26-29. (megjelenés alatt)
4. **Leidecker, Eleonóra – Galambosné Tiszberger, Mónika – Kráncz, János:** *Ízületi fájdalom gyakorisága különböző fizikai aktivitású populációban.* A Magyar Gyógytornászok Társasága VIII. Kongresszusa, Pécs, 2011. október 20-22. (megjelenés alatt)